FAR: A Computational Model for Solving Visual Intelligence Tests using Fractal Reasoning

Thesis Proposal

Brian Keith McGreggor

Design and Intelligence Laboratory
School of Interactive Computing
Georgia Tech

Last Updated May 21, 2012

Advisor
Dr. Ashok Goel

Committee members

Dr. Gregory Abowd	School of Interactive Computing, Georgia Tech
Dr. Paul Corballis	Department of Psychology, University of Auckland
Dr. Irfan Essa	School of Interactive Computing, Georgia Tech
Dr, Ashok Goel	School of Interactive Computing, Georgia Tech
Dr. Andrea Thomaz	School of Interactive Computing, Georgia Tech

Table of Contents

Introduction **7**

Two Super Powers **8**

Noticing Novelty *8*

Shifting Abstraction *10*

Problem & Research *12*

The Problem Statement *12*

The Research Question *12*

Constructing a Received World **13**

Requirements *14*

Representations *14*

Vision, and visual reasoning *15*

Challenges **16**

The Challenge of Complexity in Representation *17*

The Challenge of the Plethora of Domains *18*

The Challenge of Visual Reasoning, Itself *19*

The Challenge of Correspondence *20*

The Challenge of Models and Architectures *21*

The Challenge of Judging a Model *22*

The Challenge of Judging an Architecture *23*

Limitations **24**

Limitation 1: Commitment to a representation *25*

Limitation 2: Commitment to specific domains *26*

Thesis & Hypotheses **27**

The Thesis Statement *27*

Hypothesis 1 .... *28*

Method 28

Evaluation 28

Hypothesis 2 .... *29*

Method 29

Evaluation 29

Hypothesis 3 .... *31*

Method 31

Evaluation 31

CONTRIBUTIONS 33

SCHEDULE OF WORK 34

The satisfaction of Hypothesis 2 .... *35*

The satisfaction of Hypothesis 3 .... *36*

The satisfaction of Hypothesis 1 .... *38*

The completion of the thesis *39*

BACKGROUND AND SUPPORTING WORK 40

ON KNOWLEDGE REPRESENTATIONS 41

The roles of representation *41*

As a surrogate 41

As a set of ontological commitments 42

As a fragmentary theory of reasoning 42

As a medium for efficient computation 43

As a medium of expression 43

ON FRACTAL ENCODING AND REPRESENTATIONS 45

Fractals *46*

Fractals in the Real World *47*

The Collage Theorem *50*

The Fractal Encoding Algorithm *51*

Decomposition and Partitioning 51

Decomposition as Level of Detail 52

Searching and Encoding 53

Patterns of Searching 53

Discovering the best match 54

Photometric Correspondence 54

Exhaustive Searching and Refining Correspondence 55

Affine and Similitude Transformations 56

Similitude Transformations 57

Only Eight Transformations 58

Translation arises from searching 59

Dilation and Fractals 59

Colorimetric contraction 59

Putting it all together: a fractal code 61

The Encoding of an Image *63*

Arbitrary selection of source 63

Arbitrary nature of the encoding 63

Resilience to defects 63

Defects in sources 64

Defects in partitioning 64

Algorithmic Complexity 65

Summary of Encoding *65*

On Fractal Analogies **66**

The Central Analogy *66*

Implementing Analogy by Recall *68*

Analogy by Recall: Preparatory Stage 68

Analogy by Recall: Indexing Stage 69

Analogy by Recall: Retrieval Stage 69

Determining Fractal Features **71**

On The Ravens Progressive Matrices **72**

Tackling Ravens with Fractals *75*

Reconciling Multiple Analogical Relationships *75*

Ambiguity *76*

Refinement strategy *76*

Mutuality 76

Extending Mutuality 77

Fractal Refinement 77

Results *78*

Fractal Performance 78

On The Odd One Out **80**

Finding the Odd One Out, Fractally *81*

The segmentation phase 81

The representation phase. 81

Extended Mutuality 82

The reasoning phase 82

Reconciling Multiple Analogical Relationships 83

Relationship Space 83

Treating Maximal Similarity as Distance 84

Distribution of Similarity 84

Ambiguity....84

Refinement strategy 85

Example....*86*

Results, Preliminary Analysis and Discussion....*88*

PAPERS, POSTERS, AND PUBLICATIONS **91**

REFERENCES **93**

Introduction

IF you, gentle reader, are a neuro-typical human, it is a certainty that among the many qualities you possess, you are an expert at visual reasoning. From birth, you receive a complex visual world, and you interpret it. Diagrams and figures, landscapes and abstracts and faces, yield to your own superimposed understanding. The point of view you assume affects your interpretation of your visual system's signal, while your interpretation of this extraordinary and mundane signal causes you to modify your stance.

"The world is full of magic things, patiently waiting for our senses to grow sharper."
— William Butler Yeats

WHAT you see affects how you think, and what you think in turn affects how you see.

WHILE the act of visual reasoning may be familiar, it is fair to say that the details with which you contend while reasoning visually pose a number of questions. What is it to reason visually? Does visual reasoning impose constraints which are absent in other forms of reasoning? What are the fundamental processes of visual reasoning?

FOR such a commonplace activity, perhaps our most frequent act of creativity, visual reasoning is opaque.

THE thesis proposal you're holding is my effort to illuminate a slim path into the realm of visual reasoning. First, after a few brief opening remarks, I'll develop the problem statement, the research question, and what it is to construct a represented world. I'll note the several challenges to undertaking this work, and to the way in which I'll limit its scope. Next, I'll present my thesis statement, and the three hypotheses that this thesis will address. Finally, I believe my work will make two significant contributions to science, and I'll talk about them.

ONWARD.

Two Super Powers

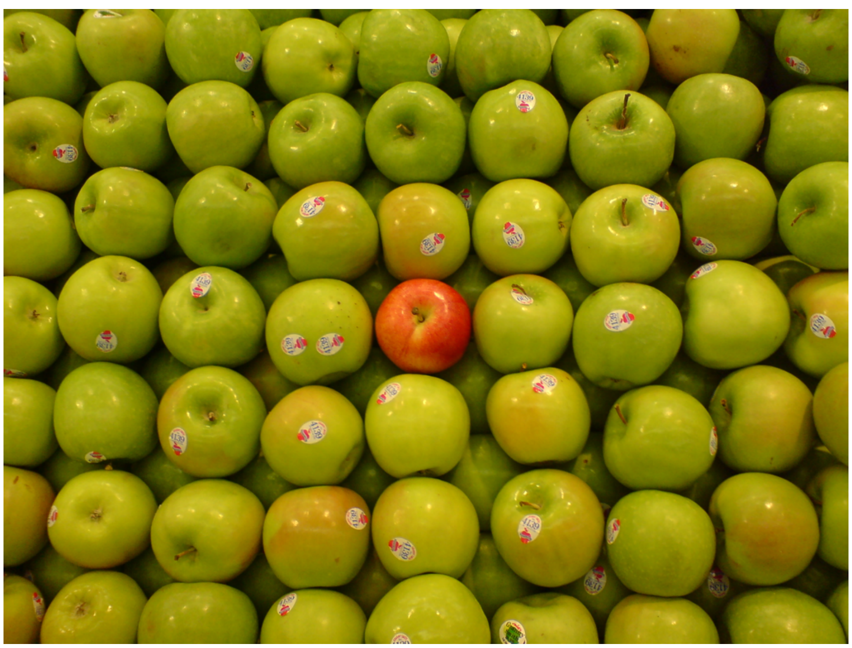
AMONG the variety of processes, there are two aspects of human visual reasoning that seem especially powerful - two super powers, if you will. These are the ability to notice novelty, and the ability to shift to an appropriate level of abstraction.

Noticing Novelty

PERHAPS as a child, you would engage, as I did, in that yearly ritual known as the Easter Egg hunt. How straightforward a task, to find those eggs hidden among the grasses and bushes. Yet, how remarkably difficult would it be to program a robot (even a remarkable one) to hunt eggs! Where might you begin, to describe the task? The incoming visual signal, the messy and ever-shifting scene, confounded by the whirlwind of other children (or other agents!) engaged in their own hunt, would be very complex indeed. Still, from this chaotic signal, children are able quite readily to pick out the elusive eggs. They notice the novelty.

To note novelty would seem to be an ability emergent from some evolutionary drive. Without the skill of novelty appreciation, our ancestors might never have distinguished bad fruit from good, or seen that predator in the distant grassland, or known their mate's display.

To me, novelty and familiarity are related and intertwined ideas–almost, but not quite purely opposite sides of the same mental coin. For you might be very familiar with some visual object, yet you may not consider it to be novel unless you encounter that object at time when you least expect it. Novelty implies a context in which you appraise the visual signal; familiarity does not necessarily suggest this. You may be entirely familiar, as an example, with what an apple looks like, but that apple would be unremarkable and lack novelty without some context.



IT is in the implicit contextualization of novelty that we find the bridge to that central core of our cognition, our ability to make analogies. Our experiences provide a rich, ever changing context in which to situate, to compare, and to remember the in-falling visual world. This textural lexicon is the structure unto which we lay the newly arriving world for judgment. Something we regard as familiar (or rather, similar, or analogous) must agree, in some sufficient number of aspects or ways or degrees, to that expectant tapestry. For something to be novel, though, we need only note a single aspect or way or degree that doesn't match.

Finding similarities is one method by which we make analogies; noticing novelty is a way by which we break them, and make newer ones.

Shifting Abstraction

THAT we are able to notice novelty so quickly, to zone in on just that substantive difference, is remarkable. How is it, then, that we are able to make such swift shifts, and draw our attention to those aspects of the visual world?

Cognitive psychology offers at least two models for how visual attention shifts.[refs] The spotlight model describes attention as having a focus area of very high visual resolution, and a fringe area surrounding the focus but with a substantially lower visual resolution. The size of the spotlight, and the relative proportion sizes of the focus and fringe, are fixed. The zoom-lens model is the spotlight model, but relaxes the constraint of the sizes. The tradeoff between these models rests in how much information is carried into the incoming signal, through the shifting in size of the region of high visual resolution. Both models maintain that the center of attention is wherever the geometric center of the focus area happens to land within the visual field.

It's that last bit–the geometric center of the focus area– that poses an issue. How do we decide *where* to focus? Perhaps we're driven by some innate properties of objects in the visual field to focus on those areas; maybe what we're thinking at the time directs our eyes to focus other places. It may be a bit of both.

Stimulus-driven, "bottom-up" processing is called exogenous attention; goal-driven "top-down" processing of attention is called endogenous or executive attention.[refs]

Even so, it seems that we regard the entire image somehow, and then some further processing happens which directs our attention to focus on certain regions. Regardless of *where* we focus, what happens is that that area of focus dominates the visual signal's information *content*. The visual field can be reinterpreted at a finer granularity if we choose. The degree to which we abstract the visual field into finer or coarser granularity (or, better, resolution) is mediated by the attention mechanism.

We effortlessly shift these levels of abstraction, changing the way in which we modulate the inbound visual world, all in the context and service of the task at hand. If the task is notice novelty, then these shifts of abstraction effectively aid and guide the hunt for just that one aspect that bestows the label *novel* to the object.

WHAT you see affects how you think, and what you think in turn affects how you see.

Problem & Research

THE challenge would seem to distill such worldly observations into the tractable.

The Problem Statement

GIVEN that such twin powers are so fundamental to your (and my, and every human's) visual reasoning, the problem lies in precisely how this may occur. In other words:

> *How do we receive this complex visual world and notice novelty at the appropriate level of abstraction?*

The Research Question

AS my work is focused on the creation of computational models, and not expressly upon the delivery of a cognitively-plausible explanation of the phenomena, my research question by necessity must be a substantial restriction of the loftier problem stated above. However, such a restriction should not be any less bold, and this suggests that I place that boldness in the choice of problem domain.

THEREFORE, I offer this as the research question to which my doctoral dissertation is addressed:

> *How might a cognitively-inspired computational model solve problems of visual similarity and novelty, such as those found on intelligence tests?*

I now shall explicate the problem and domain, and thereby motivate the thesis statement and its attendant hypotheses.

Constructing a Received World

IT is important to draw a distinction between what the world *is* and what the world *affords*. Some object in the world may be labelled as *novel* by a particular observer, but that is not sufficient to suggest that the object in question would be labelled as novel by every conceivable observer. Remember, novelty depends upon context, and every observer's context–her internal, mental context–will vary. Similarly, while the world is continuous, it does not directly offer a notion of abstraction, merely offering an opportunity for an observer to receive the world in differing manners through some enaction of the observer upon or within the world (changing the nature of the light which falls upon an object or manipulating the object somehow) or through some modification of the observer as an entity within the world (moving closer or further to an

"This world falls on me, with hopes of immortality."

- The Indigo Girls

object, or changing the visual system mechanically via squinting, and the like).

Requirements

The acts of noting novelty and shifting abstraction are cognitive acts which occur entirely within the mind of the observer. The world affords them, but it is the observer performs them. That is, some set of cognitive processes occurs within the observer to accomplish these feats. A goal of my thesis to create one or more cognitive models which contain these processes, and characterize their behavior on certain tasks, contrasting that behavior with human behavior.

These mental acts are available to be performed because the mind somehow must possess a sufficient representation of the received world which affords them.

Representations

While the term "representation" is quite commonplace in the literature and its use may be familiar to you, very rarely is tackled the notion of what a representation may be. However, a 1993 paper by Davis, Shrobe, and Szolovits, addressed this issue head-on (Davis, 1993). In their paper, Davis, et al., note that representations play five distinct, critical roles:

- as a surrogate;
- as a set of ontological commitments;
- as a fragmentary theory of reasoning;
- as a medium for efficient computation; and
- as a medium of expression.

Each of these aspects matters when regarding visual reasoning. The fidelity of the correspondence between the representation as surrogate and the received world affects and informs the possible levels of abstraction. The ontological commitment of what within the received signal to represent (and what to leave out) contribute to the constraints the representation may impose upon reasoning.

The fragmentary reasoning that a representation affords stems from what inferencing it allows, and how that set of allowed inferences may be constrained. The guidance a representation gives for computation arises from its role as an organizational mechanism for the corresponding received information, and reflects upon the adequacy with which that information is captured. The utility of the representation for communicating information directly affects our ability to mix new data with old into newer data, and provides the way in which comparison arises.

I shall direct the intrigued reader to a subsequent section, On Knowledge Representation, where I develop each of these aspects of knowledge representations in some detail, within the context of visual reasoning.

Vision, and visual reasoning

IT may be tempting to view these remarks, and indeed this entire thesis, as focused on vision. While there has been substantial research on the detection of novelty in computer vision, my efforts concern themselves with visual reasoning, and in particular the role analogy-making may play when reasoning about visual stimuli. My thesis is about cognitive strategies, and how they arise from the choices we make when representing the complex world we receive.

Challenges

The grand task is to find and describe these cognitive processes which can distinguish between the familiar and the novel, and which can shift between levels of abstraction automatically. These cognitive processes are afforded and sanctioned by some appropriate representation of the received complex visual world which affords re-representation to varying levels of abstraction and offers features which may be used for memorization, recall, and comparison. The acts of discovering and characterizing those processes are thereby co-mingled with the act of describing a suitable representation.

In this endeavor, I face several specific challenges, but each offers its own hope for resolution, either directly or by acting as a limiting guide for my research. Those challenges are:

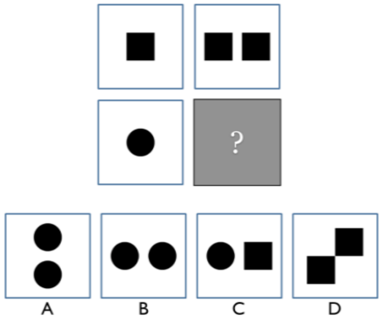
- Complexity in Representation
- A Plethora of Domains
- Visual Reasoning Itself
- Correspondence
- Models versus Architectures
- Judging a Model
- Judging an Architecture

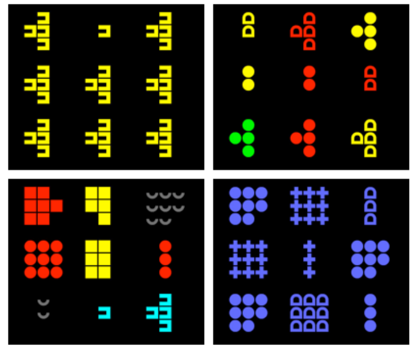
I'll describe each of these challenges in turn, next.

The Challenge of Complexity in Representation

THE world we inhabit is profoundly messy, and the visual signal we receive from it is complex. The representation must be able to capture the complexity of the received world. However, to demonstrate that the attempt to characterize and capture every conceivable aspect of the world with sufficient complexity seems foolhardy. Nonetheless, we must suppose that an example of a representational form should exist, seek to discover that example, focus our efforts therein, and extrapolate.

This challenge works for me in two important ways. It constrains my research work to a subset of the world, but it simultaneously forces me to consider a universal representation, a substrate upon which the world may be built.





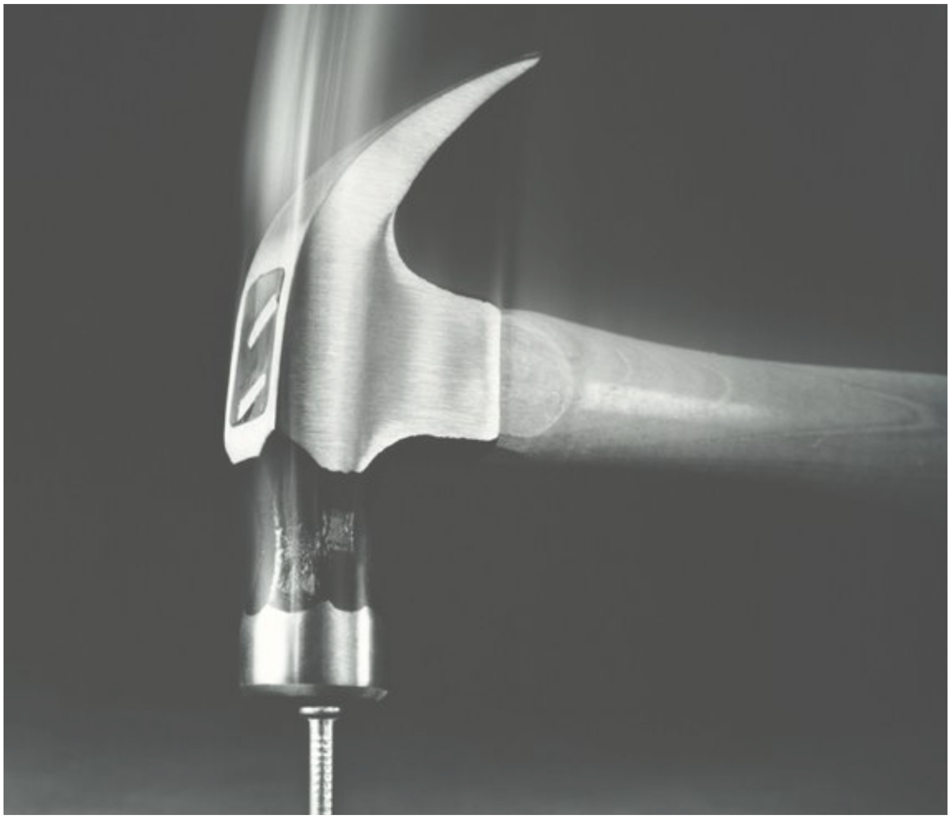
The Challenge of the Plethora of Domains

THIS is a corollary to Challenge 1. The complexity of the world offers up quite the variety of problems and puzzles. While our minds might capably perform many, many tasks, in this research quite particularly I am exploring novelty and abstraction tasks. Moreover, I am exploring them in the context of visual reasoning specifically.

Thus, I shall restrict my consideration of problems to receive from the world to those domains in which novelty or similarity may be determined via visual input alone.

Broadly (but I admit, happily), the term which best characterizes the problem domain chosen is *visual analogy*. As chance would have it, much if not most all of the prior research in visual analogy has been limited to a narrow set of problems. Generally, these problem domains have been in the area of computational psychometrics.

While it is somewhat daunting to create models and write code which will be compared against others code, and it is certainly true that one's model and code must achieve a certain measure of correctness on those psychometric tests in order to be taken seriously in literature reviews, selecting problems from psychometrics has a distinct advantage over other domains: the general breadth and availability of human performance data on those very tests.

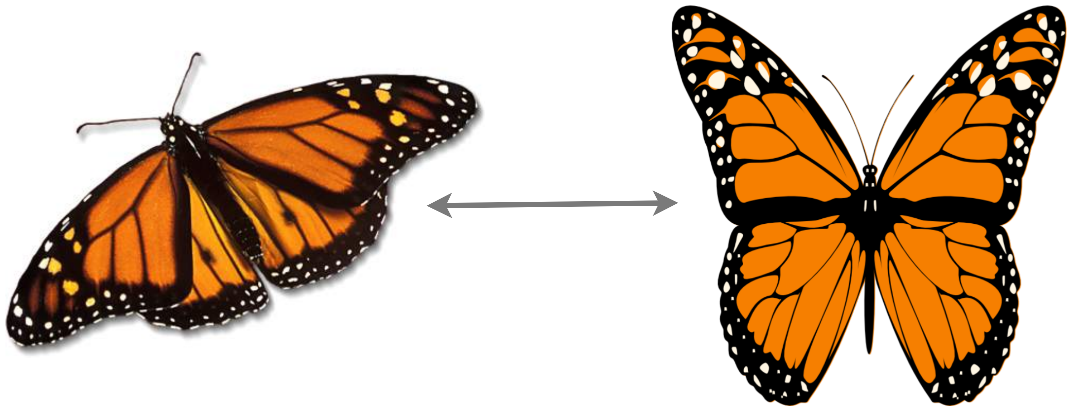


The Challenge of Visual Reasoning, Itself

IN artificial intelligence, we build models of cognition and computational creativity, and subject those models to various tests. Often, these tests are themselves artificial, contrived to limit the model's domain to a carefully composed world. Yet, a criticism of much of AI is that the composition of the problem domain is too carefully constrained, and that the resulting model clearly should work, for it, and the world upon which it acts, are joined one to another, representationally intertwined.

We can presume that there may be many different ways in which a problem may be represented. However, a chosen representation expressly determines the nature of the reasoning which may operate upon the representation, and our selection of representation must expressly afford and sanction the kinds of visual reasoning we wish to explore.

Thus, the selection of representation must be restricted to those which both affords reasoning about novelty and similarity, and supports shifting levels of abstraction.



The Challenge of Correspondence

THE current theories of visual analogical reasoning depend upon a significant theoretic leap: that the received world is transformed from a series of received percepts into some symbolic representation. The challenge is that this transduction of perception into symbolism readily can be viewed as reducing correspondence with the world (that is, with reality). Reducing correspondence with reality affects the correspondence in level of abstraction which might be afforded by the symbolic representation.

Thus, a suitable representation for my investigation must maintain as strong as practical a correspondence to the received percepts.

The Challenge of Models and Architectures

THERE exist several visual analogy problems, and my research work will address certain of these. These problems share many common aspects, but they have very specific differences as well. Each of these problems may lead to its own cognitive model. Indeed, this is the *desired* outcome.

Can we presume that a common representation can be shared amongst those models, and provide an account for their commonality? If so, then I may find that though there can be differences in model, there may exist a single cognitive architecture upon which those models are founded.

Finding first these models, and then an underlying architecture, is a goal of my work.

The Challenge of Judging a Model

COGNITIVE models are judged by their degree of empirical coverage and their parsimony. Judging parsimony is straightforward: I propose that we might note the number of computational methods needed to address the problem. Judging empirical coverage, on the other hand, is complicated.

Ordinarily, empirical coverage means that human performance levels are achieved, both in the time taken to perform a task, and in the number and kind of errors made during a task. It's a bit foolhardy, I believe, to judge a cognitive model's coverage based on time performance, for two reasons:

> *Iron changes everything* Each year, our machines grow faster and faster. At some point, the task that satisfactorily covers human performance will be performed much quicker by machine.
>
> *Exotic computation* The algorithms we design are generally executed in a serial fashion, with strict data flows. Our brains don't quite seem to follow either aspect, being inherently (and massively) parallel, and with bidirectional information flowing.

For these reasons, I hope that you will judge my models' empirical coverage by its error patterns vis-a-vis humans.

The Challenge of Judging an Architecture

COGNITIVE architectures should be judged by their unification of supported cognitive models, their degrees of freedom, and their world view.

Certainly, I have as a goal that an architecture can be extracted from the cognitive models which I develop from the problem domains. In this regard, the architecture somewhat trivially can be said to act as a unifying force across those domains. A way to perform that extrapolation from model to architecture is to seek out patterns of computation. A pattern would be akin to an algorithm, but more of a meta-description of a class of algorithms than a particular algorithm per se.

An architecture is a canvas, upon which the constituent set of processes operate, in well-delineated ways, upon clearly committed, deliberate structures. The cognitive models are built upon this canvas, and affected and sanctioned by it.

Two dangers of architectural construction are overunification and degrees of freedom. Unification seeks to make broad commentary about the nature of the cognition and computation permitted. Overunification would drive too many domain-specific techniques into the architectural substrate, minimizing the number of cognitive models which can be based upon it. The common language of the architecture, the design patterns therein, allow cognitive models to be themselves unified.

The problem of degrees of freedom occurs when a problem domain (or set of domains) is too regular, or when a particular problem domain requires too many ad-hoc techniques. I believe I can mitigate this danger by carefully selecting the problem domains (and I outline these subsequently).

Lastly, I want to describe an architecture which will afford both prospective and retrospective ways to consider other visual analogy problems. In this way, the architecture will offer a way to view cognition, and in this thesis's case, visual analogy.

Limitations

THE challenges I've just enumerated offer ways to constrain the work ahead. I hereby propose to limit the scope of my thesis in two important ways.

I am making a strong commitment to a particular kind of representation.

I am focusing my efforts on developing cognitively-inspired computational models for four specific, interrelated problem domains.

HERE is my rationale, for each.

Limitation 1: Commitment to a representation

REPRESENTATIONS are at the very core of reasoning, and I make this limitation very strongly. The representation I choose for my work is the fractal representation, which I have developed over the course of my preliminary work.

THE fractal representation arises from fractal encoding of visual input. Fractal encoding is an encoding of both spatial and photometric relationships. This encoding captures the nuances of textures present within a received image. Fractal representations capture the similarity between visual images, even if the images are the same. You will find a thorough discussion of fractal encoding and representation later in this document, in the section On Fractal Encoding and Representation.

Limitation 2: Commitment to specific domains

I will develop cognitively-inspired computational models based on fractal representations across four specific problem domains, two of visual similarity and two of visual novelty.

FOR visual similarity, I choose the Ravens tests, Standard and Advanced, and the Miller's test, used by Evans in his seminal early work on analogy. The Ravens tests offer a combined set of 108 well-documented, human-tested problems. Miller's test offers 20 problems.

FOR visual novelty, I choose the Odd One Out problem set of almost 3,000 problems, and the Dehaene test of core geometry, 45 problems.

MY intention is that by considering problems of similarity (Ravens and Millers) independently from problems of novelty (Odd One Out and Dehaene), distinct cognitive models will emerge, one for similarity and one for novelty. From these models, I will extract those domain-generic techniques to form a cognitive architecture, one in which noting novelty and adjusting levels of abstraction are fundamental and strategic acts, afforded expressly by the fractal representation.

I also point out that these problem domains are static worlds. If this thesis were concerning itself with vision in the general sense, I would have to choose additional dynamic domains which would offer the opportunity to address the challenges of occlusion, motion, noise, and the like. Although I believe my work may hold promise in those areas, this thesis is not about vision, it is about visual reasoning, and the role analogy-making and representation play in it. Thus, I am able to make these domain restrictions without loss of generality.

Thesis & Hypotheses

WITH these limitations and intentions in mind, I now can declare a sufficient, expressive thesis statement and collection of hypotheses.

The Thesis Statement

FOR this dissertation, I make the following thesis statement:

> *Fractal reasoning is a novel, feasible and useful computational technique for solving problems of visual similarity and novelty found on intelligence tests.*

MY thesis addresses the following hypotheses:

1. that using the fractal representation, a robust cognitively-inspired computational strategy may be determined which systematically adjusts to an appropriate level of abstraction;
2. that using the fractal representation, a robust cognitively-inspired computational model can be derived for certain classes of problems of visual similarity, such as the Raven's Progressive Matrices tests;
3. and, that using the fractal representation, a robust cognitively-inspired computational model can be derived for certain classes of problems of visual novelty, such as those in the Odd One Out set.

I'll now develop each of these hypotheses in some detail.

Hypothesis 1

Using the fractal representation, a robust cognitively-inspired computational strategy may be determined which systematically adjusts to an appropriate level of abstraction.

Method

I will develop and implement an algorithm by which the ambiguity or uncertainty with which an answer to a visual analogy problem by be characterized may be mapped to those features naturally arising from the fractal representations used in attempting the problem's solution. I will show that such a characterization can be used concurrent with problem solution, as a mechanism for driving level-of-abstraction refinement. I will present an analysis of the algorithms's performance, in terms of computational complexity and runtime complexity. Finally, I will propose that the reasoning embodied in the algorithm may be construed as a model of visual abstraction.

Evaluation

I claim that successfully addressing this hypothesis can be accomplished by providing crisp answers to these questions:

- Does the analysis convincingly illustrate the complexity? Does the analysis naturally and solely follow from the fractal representation?
- Does the algorithm work well enough concurrently to offer a distinct advantage to solving problems of visual analogy?
- Does the algorithm truly offer a model of visual abstraction, rendered solely from the fractal representation of the problems?

Hypothesis 2

Using the fractal representation, a robust cognitively-inspired computational model can be derived for certain classes of problems of visual similarity, such as the Raven's Progressive Matrices tests.

Method

I will describe the problems of the Raven's Progressive Matrices tests, in terms of their individual nature as well as their importance in the realm of human psychometrics. I will develop both a visual reasoning strategy and an algorithm which embodies that strategy, based upon and relying solely upon the fractal representation of a Raven's problem, which I shall demonstrate will solve the problem without intervention. Finally, I shall cause the algorithm to be implemented in code and executed against the full set of Raven's Progressive Matrices tests (the Standard and the Advanced sets), and report the results of the algorithm's performance. Similarly, I will do the same for the problems of the Miller's analogy test, with no modification to the underlying algorithm or representation.

Evaluation

Proving this hypothesis requires that I answer these questions:

- How does the algorithm perform on the Raven's Progressive Matrices tests, and why?
- What is the relationship of the algorithm's performance to previous computational approaches to the Raven's?
- For those problems which are successfully addressed, why does the algorithm succeed?
- For those problems which are not solved successfully, why does the algorithm fail?
- What do the successes and the failures of the algorithm have to say about the nature of visual reasoning upon the Raven's test?
- Did the same algorithm and representation work for the Miller's problems?
- What are the extractable aspects of visual reasoning, vis-a-vis similarity and abstraction?

Hypothesis 3

Using the fractal representation, a robust cognitively-inspired computational model can be derived for certain classes of problems of visual novelty, such as those in the Odd One Out set.

Method

I will describe the problems of the Odd One Out tests, in terms of their individual nature as well as their distinction from visual reasoning as required for problems of the Raven's test. I will develop both a visual reasoning strategy and an algorithm which embodies that strategy, based upon and relying solely upon the fractal representation of an Odd One Out problem, which I shall demonstrate will solve the problem without intervention. Finally, I shall cause the algorithm to be written in code and executed against a large corpus of Odd One Out problems (approximately 3,000, at varying levels of human difficulty), and report the results of the algorithm's performance. Similarly, I shall direct the algorithm and representation developed for the Odd One Out to address those problems present in the Dehaene set of core geometry.

Evaluation

Beyond the construction and successful execution of the code implied, these questions must be answered:

- How does the algorithm perform on the Odd One Out problems, and why?
- What is the distinction between Raven's problems and Odd One Out problems, and why is that important to illustrate the robustness of the fractal representation?
- As with Hypothesis 3, for those problems which are successfully addressed, why does the algorithm succeed, and for those problems which are not solved successfully, why does the algorithm fail?
- What do the successes and the failures of the algorithm have to say about the nature of visual reasoning upon these problems?
- What may be extracted, vis-a-vis noticing novelty and abstraction?

Contributions

MY body of research makes two primary, novel, and significant contributions to science.

THE first contribution is a parsimonious cognitively-inspired computational architecture for visual reasoning which automatically adjusts to an appropriate level of abstraction suitable to meet the demands of a variety of visual analogy problems.

THE second contribution is the fractal representation itself, a new and novel knowledge representation that will open the door for analogy researchers, cognitive scientists, and computer scientists to explore the role self-similarity and perceptual complexity play in analogy making.

Schedule of Work

I have already completed a substantial body of research on the issues presented here. There remains quite a bit of work yet to be done, and I shall put myself to the task, but not quite in strict hypothesis order, for reasons I shall explain.

HERE, then, is a rough draft of the work schedule I intend to pursue, in the order in which I expect to prosecute it. Even though the presentation below is serial, I will be working on the first three tasks in parallel, and leave the last task to begin in earnest upon their completion.

The satisfaction of Hypothesis 2

Expected completion: Fall, 2012

MY second hypothesis concerns the application of fractal representations to problems of visual similarity. Here, I will balance my time between writing and refactoring code, and writing prose.

I, along with my colleague Maithilee Kunda, already have done quite a bit of work on the Raven's Progressive Matrices test, and in particular on the Standard Progressive Matrices test (SPM). Very recently, I have turned my attention to the Advanced Progressive Matrices test (APM), and have made an initial run of the algorithm against that problem set. The results are yet preliminary, but they are very promising.

For this effort, I will be revising the existing algorithm, and bringing to it a visual reasoning strategy based upon the chief thrusts of this thesis (adjustment of level of abstraction and relationship). Once in place, I will rerun the new algorithm against the SPM and APM tests, without intervention. This code execution will take quite a bit of time (in the current implementation of the code, each fractal encoding takes between 10 and 20 seconds of a single core's processing power, on a contemporary multi-core machine, and there are more than 200 individual fractals to calculate in each 3x3 Raven's problem for each level of resolution). Fortunately, the results of the fractal encoding are cached, and the subsequent algorithm execution which solves a Raven's problem takes very few seconds. Rapid iteration will allow me to refine the strategy over a comparatively brief time period (less than two weeks).

Once the algorithmic iteration is accomplished, I will run the strategy against every available Ravens test (there are several other variants beyond the SPM and APM). I will conduct a thorough analysis of the results of these executions, and characterize the performance against human contemporaries. I also intend to run the strategy against for the problems of the Miller's analogy test, and make a similar analysis and comparison.

It is my expectation that the strategy will perform well against some of the problems, and not so well against other problems. Therefore, a portion of this task is to classify, to some extent, which problems belong to each set, and to seek those characteristics which suggest a priori classification. This characterization of the problems will further inform the model, and especially the architectural considerations which play a key role in the work on Hypothesis 1.

In particular, I will be exploring the role of directionality in the relationships (I explain this concept below in the supplemental discussion of fractal representation), as the problems of the Raven's test appear to have a strong left-to-right and top-to-bottom production preference. This belief is supported by some of the remarks in the Raven reference material on errors made by human test takers. A discovery of directional characteristic would allow for predictions to be made with respect to the model's performance on other like problems, and would allow me to make a stronger case for empirical coverage.

The satisfaction of Hypothesis 3

Expected completion: Fall, 2012

MY third hypothesis concerns the application of fractal representations to problems of visual novelty. Just as my plan for satisfying Hypothesis 2, I will balance my time between writing and refactoring code, and writing prose.

Here also I have done much prior work, especially on the Odd One Out problem set. I have already run an initial version of the model against all problems at my disposal, and have presented the preliminary analysis of this work at the recent Fall AAAI Symposium on Cognitive Systems.

In support of this thesis, I will revisit my analysis of the algorithm. It seems apparent from my initial work that the problems fall into certain bands of difficulty, and the strategy I have used thus far fairs less well against certain classes of those problems. Thus, I believe there may be some distinction in these categories which may be determined. In examining these distinctions sufficiently, I seek to find at what point a confluence of factors might reach a complexity which cannot be easily ferreted out by the strategy. I'll revise the code, guided by the insights gathered, and rerun the strategy against the entire lot.

The self-segmentation of the problems into bands also will guide my model extrapolation into the architecture for Hypothesis 1. One of the avenues that I will be exploring here in particular is the interplay between the resolution of the partitioning and the size of the relationships under consideration. It is possible that the banding present in the preliminary work is only due to the use of pairs of figures, and not triplets or higher degrees. Thus, there may be bands or classes of problems of novelty for which a solution may be attained only if relationships greater than pairs are considered. A careful examination of these problems may yield visual characteristics which are encoded by the representation, yet escape merit. By contrasting these sets, beyond merely classifying the problems, I will be seeking the hallmarks of feature saliency.

As with the Ravens code activity for Hypothesis 2, the creation of fractal representations of the relationships present in these problems takes quite a bit of time, for there are 72 fractals to calculate per each problem, at between 10 to 20 seconds per fractal, at each level of abstraction. Thus, for the 3,000 problems, at five levels of abstraction, will take perhaps as much as 6,000 hours of computation time. Fortunately, the creation of the fractal representations can run in parallel, across multiple cores on multiple machines, and is independent of the strategy which uses them. The strategy's action upon the representations takes only a handful of seconds per problem, and this will permit rapid iteration of the strategy in an analogous fashion to the evolution of the strategy for the Raven's tests.

I will round out this section of work by complementing the Odd One Out analysis with an analysis of the strategy's results against those problems present in the Dehaene set of core geometry. This set of problems is also likely to yield insight into which factors may be discernible through fractal representation.

The satisfaction of Hypothesis 1

Expected completion: December, 2012

MY first hypothesis forms the basis of the first of the two contributions: a parsimonious cognitively-inspired computational architecture for visual reasoning which automatically adjusts to an appropriate level of abstraction suitable to meet the demands of a variety of visual analogy problems. The principal action for me during this task is to write, and not to code. Therefore, while I may direct some efforts toward this hypothesis during my prosecution of satisfying the other hypotheses, the work I have already done and the insights I am to gain while pursuing those hypotheses will have direct bearing on this architecture, and the balance of this effort must follow those.

From careful analysis of the characterization of the problems of visual similarity and novelty, and of the results of the models I employ against those problems, I will derive an architecture. My intention at the outset is to illustrate the architecture as a staged approach, beginning first with the representation of the image constituents in a fractal fashion, and then recruiting or inhibiting from a class of techniques those suggested by the task at hand.

I will build upon (and reinforce and inform) the second contribution of my thesis–the fractal representation–by constructing an argument that the architecture and the reasoning therein is sanctioned by the representation, that the ontological commitments of the representation are made manifest by the architecture, and the architecture offers an instance of how the representation operates as a medium for effective communication. The ontological commitments, in particular, are important to draw out in an architectural sense, for they are the way in which the fractal representation places emphasis on the parts of the world which are relevant.

During this phase of the work, I will develop an analysis of the architecture (and of the subsequent models of visual similarity and visual novelty) performance and computational and runtime complexity. I intend to do this to be complete. I also intend this as a stage setting, not for the purposes of this thesis, but as a way perhaps to make predictions for future experiments or to act as a lens with which to regard prior experiments. In particular, the characterization of problems with regard to their directionality (in the case of visual similarity) or relationship cardinality (in the case of visual novelty) would afford a possible class of predictions and test design for exploring their appearance in human test takers. Therein may prove a way for the derived architecture to comment on human cognition.

The completion of the thesis

Expected completion: December, 2012

THROUGHOUT all the above tasks, I will be writing the various chapters and fragments of the thesis. The document you hold now is a beginning, a zeroth version of the final.

THE computational aspects of the tasks are large. However, over the course of the last few years, I have already developed a solid code framework from which to begin: indeed, I have already begun in several cases. Largest portion of the computation time lies in the calculation of the fractal representations. Fortunately, most of the earlier work has generated those representations which I may now use, and thus focus my coding efforts upon model and architecture development and refinement.

I expect to cease coding sometime in late September or early October. At that point, all of my efforts will be on the analysis of the data, and writing a coherent argument for that analysis.

THEREFORE, I anticipate standing for the defense of this thesis in the early spring of 2013.

Background and Supporting Work

I have been working on this family of research for quite some time now. Much has been written and published so far, and there are several papers in varying stages of the publication pipeline.

In the remaining sections of this proposal, I seek to provide the curious reader a deeper dive into several of the areas mentioned above. A set of published papers by myself and my collaborators is presented as well.

On Knowledge Representations

Acts of cognition involve the manipulation of knowledge, represented in some manner. While the term "representation" is quite commonplace and its use may be familiar, it is significant to note that very rarely is the notion of what a representation actually may be tackled. However, in the AI literature, a seminal paper by Davis, Shrobe, and Szolovits, took this issue head-on (Davis, 1993). In this section is a brief review of their analysis.

A terminology note Herein, I use the term **representation** as a short cut for **knowledge representation**. There may be many other interpretations of the word.

The roles of representation

A representation can be said to have meaning when in service toward a particular task. Davis, et al., note that representations play five distinct, critical roles. Those roles are as a surrogate, as a set of ontological commitments, as a fragmentary theory of reasoning, as a medium for pragmatically efficient computation, and as a medium of human expression. Let us consider each role in brief, and begin to bring aspects of visual search into the discussion.

As a surrogate

When a mind reasons about its world, this reasoning occurs internally, while the majority of what it reasons about exists externally. A representation then must act as a surrogate for things which exist outside the reasoning agency. Direct interaction with real world objects are paralleled by operations upon the internal representations of those objects.

Davis, et al., raise two significant points concerning surrogates: what is a surrogate a surrogate for, and what is the fidelity of a surrogate? Some correspondence between the surrogate and its counterpart in the world must be specified. With respect to fidelity, what attributes of the original are preserved, omitted, or implied with the surrogate must be addressed, for perfect fidelity is impossible.

Representations, then, must be imperfect, and since reasoning operates upon representations, so to must reasoning itself arrive at imperfect conclusions, even if the reasoning process itself is sound. It is this correspondence aspect which must be adequately addressed in any system which seeks to concern itself with levels of abstraction.

As a set of ontological commitments

Selecting a representation involves a decision about how and what to represent from the arriving world. A set of commitments, then, is made that both define the extent of the representation's capture of the world and define the way that extent is expressed or embodied within the representation ontologically. Here, the task at hand acts as a guide toward the selection of an appropriate ontology. These commitments start at the moment a representation begins to form, and likely accumulate as the representation is used. As Davis, et al., note, the representational power lies in the correspondence of the representation to something in the world and in the constraints that that correspondence impose.

As a fragmentary theory of reasoning

Representations are formed to allow cognition to occur within some agency. Even though the theory of reasoning arising from a representation may be implicit, it can be seen through three aspects: what the representation defines as inferencing, the set of inferences it allows, and the subset of those inferences which it recommends. The author refers the gentle reader to the Davis paper for a thorough discussion of what it is to make intelligent inferences.

Allowed inferences are those inferences which can be made from available information. As a representation might arise in any number of ways, so too might the allowed inferences vary. As Davis, et al., point out, this flexibility is acknowledged so as to admit the legitimacy of the various approaches. Having this flexibility at its core provides a framework for re-representation.

Clearly, the set of allowable inferences may become untenably large. A smaller, constrained subset of these inferences is necessary. Whether by specifying the constraints with which to select recommended inferences, or by providing them somewhat explicitly, some process or reasoning or insight must be at work to frame them. In this way, Davis, et al., citing Minsky as an example, illustrates that representation and reasoning are intertwined in a deep, theoretical manner. They also observe that much of the reasoning which informs recommended inferences has been provided by observation of human behavior.

As a medium for efficient computation

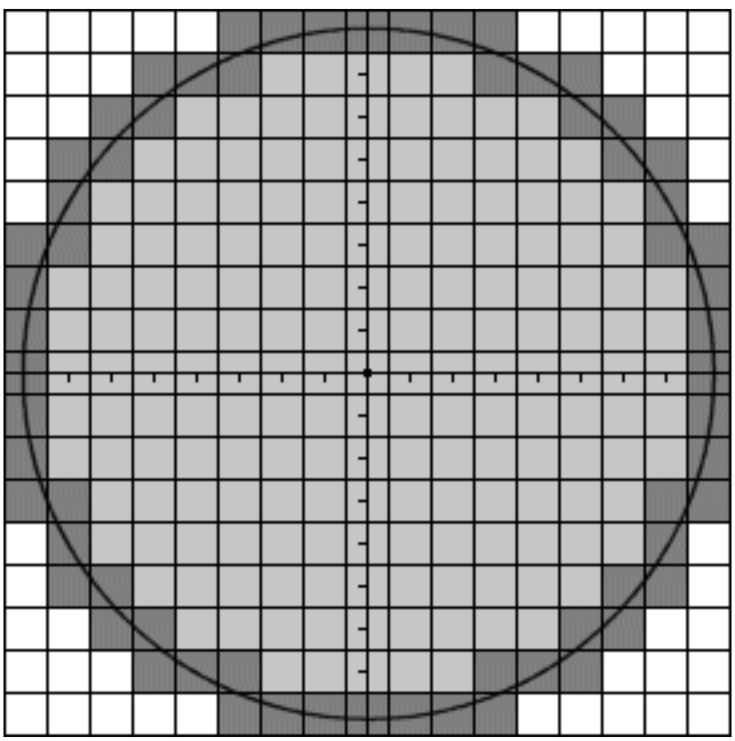
The information processing stance of human cognition holds that cognition is a computational process. In the same sense that a representation recommends inferences, so to does it imply the manner in which it may be used in computation. This guidance speaks to the adequacy of the representation, as an organizational mechanism for information, for the task at hand.

As a medium of expression

Although the Davis paper addresses itself to the notion of representations as vehicles for human expression, the author wishes to stress that the internal dialogue of, about, and with representations is as important as the external one. In so complex a system as the human brain, information must pass from subsystem to subsystem, preferentially without substantial degradation and with increasing specificity. The expression of representations internally is a process of systematic reassembly of aspects of those representations into new ones, through which other systems may operate upon the newfound representations, with the core roles of representations implied by those systems' tasks. Herein, we form cognitive models; herein, we find the basis for cognitive architectures.

On Fractal Encoding and Representations

An image, as held in memory in a computer, is a representation which may occur in a variety of forms. In one case, a vector image, the image might be represented as a proximal sum of a variety of lines, curves, and polyhedral shapes. Vector images are quite well suited for representing diagrams. In another, more common example, the image might be represented in bitmap fashion, a rectilinear array of pixels (photometric values) of a specific width and height. Bitmapped images are typically used as methods for storing so-called "natural images." In either case, a coordinate system typically is inferred to ascribe the position and orientation of various spatial elements, be they pixels or polygons.



The challenge of representing an image, in any fashion, stems from this: to what end is the representation intended? As we have seen in the previous section, a representation entails a set of possible inferences, and implicates a surrogate standing. An image representation is arrived at from some putative input. We receive the world, and we represent it.

Fractals

Benoit Mandelbrot coined the term "fractal" from the Latin adjective *fractus* and its corresponding verb (*frangere*, "to break" into irregular fragments), in response to his observation that shapes previously referred to as "grainy, hydralike, in between, pimply, pocky, ramified, seaweedy, strange, tangled, tortuous, wiggly, wispy, wrinkled, and the like" could be described by a set of compact, rigorous rules for their production [32].

The computer graphics community has generated fractal imagery, similar to this figure, for several decades.

While the formula for generating fractal imagery is quite well-known, many images of real-world artifacts appear to have "fractal" properties. If these images are "fractal" in some sense, then what formula (to be more specific, what representation) may underlie these images? What, precisely, is fractal?

Fractals in the Real World

The mathematical derivation of fractal image representation expressly depends upon the notion of real world images, i.e. images that are two dimensional and continuous (Barnsley & Hurd, 1992). Both of these assumptions are important. That an image is two dimensional means that there is an ability to assign a coordinate system to the image, and that the photometric elements, the pixels, within that image have a spatial relationship to one another (that there is a distance metric upon the space). That an image is continuous implies that no matter how closely one might choose to examine the image, there still will remain finer and finer gradations of the pixels. In a sense, the continuity of the image suggests that the selection of an image's resolution (the ability to resolve or describe a single pixel) is under the control of the observer. In this assumption, a pixel gains the descriptive quality of a photometric region.

Real world imagery, in the definition above, encompasses not only that which occurs in the natural world, but all imagery. Natural and artificial scenes, all diagrams and schemata, every image which arises as a result of light being reflected by or transmitted from any surface and subsequently falling upon the photoreceptors and made available to the human visual system is a real world image. Images generated internally or those arising from some act of visual imagination or via some other means (specifically, those images whose arrival does not encompass perception and the enactment of the early visual system, the lensing system, and especially, the striate and pre-striate cortex) are excluded from our definition of real world imagery.

A key observation by Barnsley and Hurd is that all naturally occurring images we perceive appear to have similar, repeating patterns. Another observation is that no matter how closely you examine the real world, you find instances of similar structures and repeating patterns. The twin ideas, of repeating patterns and of repetition at differing scales (or resolution), combine to provide the basis for labeling such images as "fractal." Importantly, the repetitive nature of these images persists at all observable scales, down to the resolving power of the observer.

These powerful observations suggest that it is possible to describe the real world in terms not of traditional graphical elements, but of observed similarity and repetition alone. This is the crucial idea upon which the fractal representation is formulated.

The Collage Theorem

Computationally, to determine the fractal representation of an image requires the use of the fractal encoding algorithm. This algorithm seeks to encode a given image, creating a representation of the arriving image. As we have just noted, the fractal representation is anchored by describing only observed similarity and repetition information. From what source and what manner are these similarities and repetitions noted?

At the heart of the fractal encoding algorithm is a remarkable theorem of Barnsley and Hurd, the Collage Theorem. The theorem can be stated concisely:

> For any particular real world image **D**, there exists a finite set of affine transformations **T** which, if applied repeatedly and indefinitely to any other real world image **S**, will result in the convergence of **S** into **D**.

It is clear from the prior discussion of real world imagery that a real world image may be described in terms of itself (that is, using itself as the source for noticing similarity and repetition information). The surprising principle of the collage theorem is that it does not matter what source image S is used: a set of transformations may always be found which can be guaranteed to converge upon any desired destination image D.

We now shall present the fractal encoding algorithm in detail, and illuminate the concepts contained in the collage theorem.

To determine the fractal encoding Frac(S,D) of an image D given a source image S:

Partition destination image D into a set of N images {d_1, d_2, d_3, …, d_n}. These individual images are sets of pixels.

For each image d_i:

- Examine the entire source image S for an equivalent image s_i such that a transformation of s_i will result in d_i. This transformation will be a 3x3 matrix, as the pixels within s_i and d_i under consideration can be represented as the 3D vector <x, y, c> where c is the photometric property (color) of the 2D point <x,y>. Collect all such transforms into a set of candidates C.
- Select from the set of candidates the transform that most minimally achieves its work, according to some consistent metric.
- Let T_i be the representation of the chosen transformation of s_i into d_i.

The set Frac(S,D) = {T_1, T_2, T_3, …, T_n} is the fractal encoding of the image D, given S.

The Fractal Encoding Algorithm

Given a destination image D, the fractal encoding algorithm seeks to discover a set of transformations T, such that that set of transformations, when applied repetitively to a given source image S, will result in the regeneration of the destination image. The steps for encoding an image D are shown above.

The algorithm consists of two segments: decomposition and searching. We shall examine each in turn.

Decomposition and Partitioning

The destination image D may be decomposed, or partitioned, into a set of other images. This partitioning is possible because D is given as a real world image, and is therefore continuous. Practically speaking, as D will be a computer image, there will be some effective limit as to its continuity: D will have a finite resolution, and thus a limit in the smallest achievable partition. We shall take as axiomatic that this smallest resolvable unit is a unitary pixel, and that the value of this pixel shall be interpreted as the photometric value of the region of the real world image subtended by this pixel.

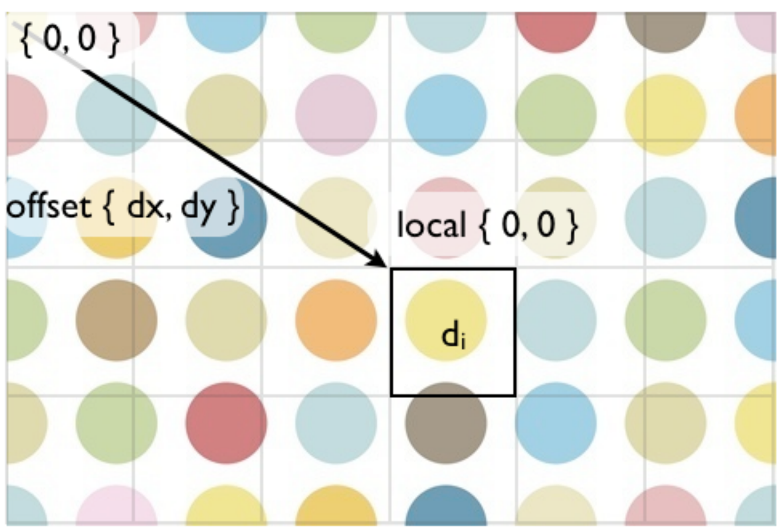
We may decompose D into other images using a variety of methods. We use region connection calculus (RCC) to impose several strictures upon the decomposition, however: each image in the decomposition must be externally connected (in a region sense, EC) to at least one other image in the decomposition, the union of all of the decomposition images must be exactly equivalent to the original image D, and the intersection of all such images must be empty. Decompositions which satisfy these criteria form a partitioning of D which is a complete, topologically closed cover over D. A straightforward manner with which to achieve a satisfactory decomposition is to impose upon D some uniform, rectilinear grid, and select the partitioning based upon some chosen grid size, as expressed in units of pixels.

Decomposition as Level of Detail

The choosing of a grid size, and of a partitioning in general, may be interpreted as a indication of the level of detail at which an image is encoded. Thus, the coarsest level of detail possible for an image is the partitioning into a single image (the whole image). The finest level of detail achievable is the partitioning of an image into that set of images wherein each image is but a single pixel (a grid size of 1). The ability to express level of detail as an artifact of partitioning, whether by controlling grid size, by altering the consistency of partition size (e.g. preferentially forming a non-uniform partitioning), or by modification of the shape and nature of the underlying regions and their spatial arrangement (i.e. hexagonal versus rectilinear scaffolding, or polar versus cartesian coordinates) is an important aspect of the encoding, and a key feature entailed by the fractal representation.

Searching and Encoding

Once the image D has been partitioned into several other images, the fractal encoding algorithm conducts a search for each image in the partition to determine which fragment of the given source image S best matches a particular partition. The method by which the search is conducted may be varied, as can the meaning of what is said to be a "best match."



Patterns of Searching

An image fragment d_i from the destination image D is a region containing some number of pixels which are addressable in some fashion. The addressability of these pixels may be viewed as a local coordinate system imposed upon the region. The region described by the fragment has a location and orientation within the destination image which is strictly determined by the partitioning. Thus, it is useful to consider the image fragment d_i as an ordered set of pixels, which have both a local coordinate system and extent, and a global position and orientation.

Discovering the best match

The source image, S, is examined to determine which fragment of it, which we shall label si, can be said to "best match" the sought for fragment di in the destination image D. That is, the correspondence between fragments si and di is found to be "best" if it is the minimum value of the following function:

$$\text{Correspondence}(s_k, d_i) = \text{PhotometricCorrespondence}(\text{Transform}(s_k, T), d_i) \quad \forall\, k \in \{ 0..n \},\ T \in \text{AdmissibleTransforms}$$

where n is the number of fragments of S which possess the same size and topological arrangement as the fragment di, and where the AdmissibleTransforms are a finite set of affine transformations.

Photometric Correspondence

The photometric correspondence between two fragments, one from the source image, and di from the destination image, is calculated to be the difference between the photometric values found in those fragments under a given alignment of those fragments. Ideally, this difference would be 0 if the two fragments were identical photometrically. An algorithm which calculates the photometric correspondence, which presumes that the two fragments may be considered as an array of such values (pixels), is given:

Start with PC = 0.

For each pixel d_{pi} in fragment d_i and its corresponding pixel s_{ki} in fragment s_k:

$$PC \mathrel{+}= (\text{Photometric}(d_{pi}) - \text{Photometric}(s_{ki}))^2$$

The value PC is then the photometric correspondence between s_k and d_i.

The Photometric value of a pixel used in this calculation may vary according to the nature of the image itself. For example, if the image is in full color, the photometric value may be a triplet of actual values; if the image is monochromatic, then the photometric value will be single valued. Since it is desired to calculate a photometric correspondence which is single-valued, a mapping from multivariate photometry to a single value is typically employed. This can be seen, globally, as mapping from one color space into another. For example, to reconcile traditional computer graphics images given in triplets of red, green, and blue values into single grayscale values, a formula such as this may be used, which seeks to equate the colorimetric luminance of the RGB image to a corresponding grayscale rendition:

$$\text{Photometric}(\langle p_{red}, p_{green}, p_{blue}\rangle) = 0.30 * p_{red} + 0.59 * p_{green} + 0.11 * p_{blue}$$

Careful consideration of the underlying photometric nature of the image being encoded therefore must be given, but only at this particular moment in the overarching algorithm for encoding. The choice of the Photometric function determines the interrelationship of the image's colorimetry and its constituent importance to the matching function.

Exhaustive Searching and Refining Correspondence

The search of the source image S for a matching fragment is exhaustive, in that each possible corresponding fragment s_k is considered regardless of its prior use in other discovered transforms. Every one of the n fragments of S which may be matched against d_i are matched and evaluated according to the correspondence formula.

We note that there may be many fragments in the source image which may have identical photometric correspondence to the sought for fragment d_i. This is particularly true when all of the values in the two fragments are identical. To break these potential ties, a further refinement of the correspondence function is necessary.

We compute a simple distance metric upon the images, and give it a weighting. Thus, the correspondence calculated between two fragments becomes:

$$\begin{aligned}\text{Correspondence}(\, s_k, d_i\,) = {} & w_1 * \text{PhotometricCorrespondence}(\, \text{Transform}(\, s_k, T\,), d_i\,) \\ & + w_2 * \text{Distance}(\, s_k, d_i\,) \\ & \forall\, k \in \{\, 0..n\,\},\ T \in \text{AdmissibleTransforms}\end{aligned}$$

where the weights w_1 and w_2 are chosen such that the calculation of correspondence is dominated by the value of the photometric correspondence. This can be ensured if the following relationship is held:

$$w_2 * \text{maximalDistance} \ll w_1 * \text{minimalJustNoticeablePhotometric}$$

where maximalDistance is the longest possible distance between the origins of d_i and any fragment in the corresponding source image, and minimalJustNoticeablePhotometric is the PhotometricCorrespondence which would be calculated if the photometric difference between d_i and any fragment were as small as possible yet not zero.

Affine and Similitude Transformations

The fractal encoding algorithm seeks to find the best matching fragment in a source image which corresponds to a given fragment from the destination image. As shown above, this matching is achieved by calculating the photometric correspondence function between two fragments, while considering all admissible transformations of the fragment from the source. The set of admissible transformations is a subset of affine transformations known as similitude transformations.

An affine transformation, in two dimensions, may be considered to be of the form:

$$W(x,y) = (ax + by + e,\ cx + dy + f\,)$$

where a, b, c, d, e, and f are all real numbers. This equation, which maps one point in a two-dimensional plane into another point in a two-dimensional plane, may be rewritten into matrix form like so:

$$W(<x,y>) = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} e \\ f \end{bmatrix}$$

In this way it can be seen that an affine transformation is a combination of a linear transformation followed by a translation.

A number of nameable, interesting transformations of fragments can be formed, with specific values used for those six real numbers:

Identity:	$W(x,y) = (x,y)$	$a = d = 1.0$, $b = c = e = f = 0$
Dilation:	$W(x,y) = (r_1 x, r_2 y)$	$a = r_1$, $d = r_2$, $b = c = e = f = 0$
Reflection:	$W(x,y) = (x,-y)$	$a = 1.0$, $d = -1.0$, $b = c = e = f = 0$
	$W(x,y) = (-x,y)$	$a = -1.0$, $d = 1.0$, $b = c = e = f = 0$
Translation:	$W(x,y) = (x+e,y+f)$	$a = d = 1.0$, $b = c = 0$

Not all affine transformations are admissible for the fractal encoding transform. In particular, those which are admissible must be invertible. Intuitively, this means that each point in space can be associated with exactly and only one other point in space. Mathematically, this means that the inverse has this form:

$$W^{-1}(x,y) = (dx - by - de + bf, -ex + ay + ce - af) / (ad - bc)$$

and the denominator $(ab\text{-}bc)$ must not be equal to zero to satisfy invertibility.

Similitude Transformations

An important group of affine transformations are those which are called similitudes. A similitude transformation may be expressed in one of these two forms:

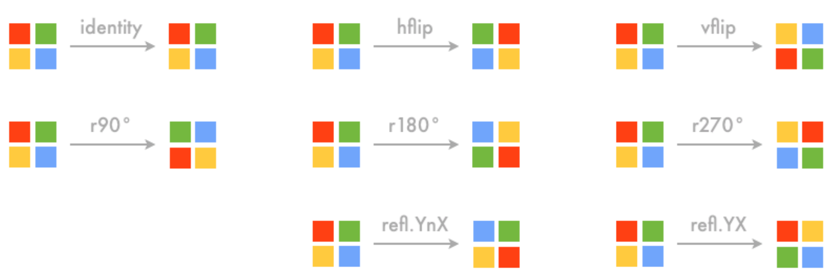
$$W(\langle x,y\rangle) = \begin{bmatrix} r\cos\Theta & -r\sin\Theta \\ r\sin\Theta & r\cos\Theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} e \\ f \end{bmatrix}$$

$$W(\langle x,y\rangle) = \begin{bmatrix} r\cos\Theta & r\sin\Theta \\ r\sin\Theta & -r\cos\Theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} e \\ f \end{bmatrix}$$

Thus, a similitude transformation is a composition of a dilation factor **r**, an orthonormal transformation (a rotation about the angle Θ where $0 \leq \Theta < 2\pi$), and a translation **(e,f)**. Similitude transformations are invertible.

Only Eight Transformations

Given this formulation for similitude transformations, one can imagine having to consider a great many potential rotational angles to find the best match. Indeed, the computational complexity of the encoding would seem a function of the angles under consideration. In practice, we find that we may limit ourselves to considering only eight of these orthonormal transformations.



Consider the smallest region of pixels for which orthonormal transformations upon those pixels would result in a visible change. The size of this region is an area two pixels wide by two pixels high. This small region has four lines of symmetry. Taking into account each line of symmetry, and reflecting the pixels in the region about each in turn, we find that there are eight possible outcomes.

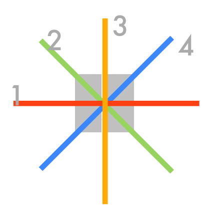
Our implementation of the fractal encoding algorithm examines each potential correspondence under each of these possible transformations, and no others. These are the admissible transformations. The transformation from this set which yields the best photometric correspondence is noted by the search algorithm.

Translation arises from searching

The searching process examines each potential fragment in a given source image for correspondence to a particular fragment of the destination image. Let us presume that we may align the coordinate systems of the source and the destination images such that their origins exactly coincide. Then, the relative location of a potential fragment in the source image can be mapped to a location within the destination image. This mapping, from the potential fragment's local origin to the particular fragment's local origin, is a translation, and it is this mapping which forms the translation portion of the sought-for affine transformation.

Dilation and Fractals

Taken together, the orthonormal transformation and the translation provide a sufficient means for describing self-similarity which may exist within an image. However, that self-similarity is not quite sufficient for describing how the similarity may occur at different levels of detail. The dilation factor, **r**, is used to invoke a contraction of space, whenever **r** < 1.0. The fractal encoding algorithm prescribes that the dilation factor to be used when searching may be conveniently set as **r** = 0.5. In practice, this entails that the source image, as a whole, may be scaled to one-half its original size, and then searched for photometrically corresponding fragments. Mathematically, choosing **r** < 1.0 ensures that the encoding derived for the entire image, if applied successively and indefinitely to an image, will cause the resulting image to converge upon the desired destination image (Barnsley & Hurd, 1992).



Colorimetric contraction

As a final step, having located the best photometrically corresponding source fragment, the algorithm determines a rate at which the two regions may be brought into colorimetric harmony. To do this, the average colorimetric description of both regions is calculated, and the distance between the two is multiplied by a dilation. The formula our present implementation uses to calculate the colorimetric contraction is:

colorimetricContraction(s_i, d_i) = 0.75 * (colorimetricMean(d_i) - colorimetricMean(s_i))

where the colorimetricMean of a region is merely the average of all colorimetric information available in that region, taking into account the multivariate nature of the underlying image as previously discussed. We have found that it is computationally advantageous to precalculate the colorimetric mean for each of the regions, in both the source and the destination images.

Putting it all together: a fractal code

For each fragment $\mathbf{d}_i$ taken from a partitioning of the destination image $\mathbf{D}$, the fractal encoding algorithm locates, via exhausive search over a source image $\mathbf{S}$, a corresponding fragment $\mathbf{s}_i$ which the algorithm has deemed to be most minimally distant photometrically under a discovered transformation. The algorithm constructs a description of its discoveries, in a representation called a fractal code. A fractal code consists of the six following elements:

Spatial	s_x, s_y	Source fragment origin
	d_x, d_y	Destination fragment origin
	T	Orthonormal transformation
	S	Size/shape of the region
Photometric	C	Colorimetric contraction
	Op	Colorimetric operation

Note that the dilation factor, for both spatial and photometric properties, is not represented here. This is for efficiency, as these dilations are presumed to be global. Further efficiencies of expression may be found by combining the two coordinate system references into a single offset vector, and by dropping the colorimetric operation (a way of describing how the colorimetric contraction value is to be combined into the region). Since the set of orthonormal transformations the search mechanism uses is finite, we may represent the transformation as a referent to that transformation's ordinal membership in the set. The size and shape of the region may be reduced itself, if the partitioning of the image is regular. In our implementation, we use a regular, uniform partitioning, which forms a grid. Thus, we can express the size and shape of the region with a single integer, which represents the width and height of the region in pixels.

Taking all of these into account, we can render a fractal code quite compactly, using five numbers:

Spatial	o_x	translation in X
	o_y	translation in Y
	t	ordinality of transformation
	g	grid size
Photometric	c	Colorimetric contraction

The entirety of the photometric information underlying a region, measuring g x g pixels, may be encoded thereby into a single fractal code, a substantial reduction.

The Encoding of an Image

The fractal encoding T(S,D) of an image then is the collected set of fractal codes determined for each fragment d_i arising from the partitioning of the image D, a set of instructions as it were, which may be said to signify the transformation of image S into image D. This set of codes is unordered, and may be applied in any sequence, so long as the entirety of the set of codes is applied before any element of the set is repeated. According to the Collage Theorem, successively and indefinitely applying the discovered fractal encoding transformations upon the fragments of S will cause S to converge into D.

Arbitrary selection of source

Note that the choice of source image S is arbitrary. Indeed, the image D can be fractally encoded in terms of itself, by substituting D for S in the algorithm. Although one might expect that this substitution would result in a trivial encoding, one in which all fractal codes correspond to an identity transform and zero translations, in practice this is not the case, for the algorithm constructs a fractal encoding of D to converge upon D.

Arbitrary nature of the encoding

The cardinality of the resulting set of fractal codes which constitute the fractal encoding is determined solely by the partitioning of the destination image. However, the ordinality of that set is arbitrary. The partitioning may be traversed in any order during the matching step of the encoding algorithm. Similarly, once discovered, the individual codes may be applied in any order at all, so long as all of the codes are applied in any particular iteration, to satisfy the constraint of the Collage Theorem.

Resilience to defects

The encoding process is resilient to defects which might be present in the source image S, and as well as to defects in the partitioning of the destination image D. Defects of these natures can occur when processing real-world imagery.

Defects in sources

A defect in a source image may be considered a portion of that image which is unavailable for comparison. Another kind of source image defect could be that some portion of the image is noisy or otherwise imperfect. These imperfections may be due to faulty reception of photometric information from the world, or an imprecise reconstruction of a desired image from memory, where the imprecision might be spatial or photometric in nature (or both).

The fractal encoding algorithm is resilient with respect to these defects in that it may take one of several choices of action:

- If the region under consideration in the source image is known to be defective, that region can be skipped by the search algorithm;
- The defective region, if known, could be used as a basis for comparison anyway; or
- Another image region could be substituted for known defective regions, where the choice of the substitution may be arbitrary.

In any case, the algorithm will encode the given partitioning of the destination image according to the information available to it during execution time.

Defects in partitioning

Just as the source image may be defective, so too might the destination image be defective, in all of the ways and for all of the reasons outlined above. The handling of defects in the destination image may be strictly determined as a decision of whether to encode known defective areas or to ignore them.

If a decision is made to ignore known defective regions within the destination image, then the cardinality of the partitioning is decreased. A reconstruction of the destination image from the representation would have apparent gaps in the imagery corresponding to the defects as noted.

If the decision is made to encode known defects, then the cardinality of the partitioning remains the same, but the resulting image that may be reconstructed from the encoding will possess the present defect.

Algorithmic Complexity

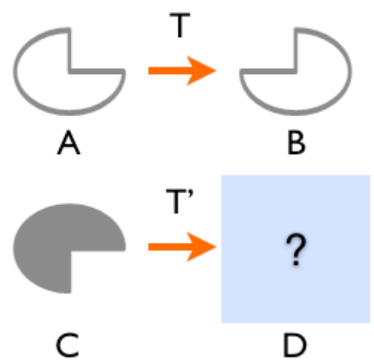
That the ordinality of the set of fractal codes is arbitrary affords another important consideration: the discovery of each fractal code may occur in parallel. This, in turn, changes the upper bound of the algorithmic complexity of fractal encoding, from $O(N^2)$, to $O(N)$.
Further implications are some codes may be found before others.

Summary of Encoding

The fractal encoding algorithm transforms a real world image into a much smaller set of fractal codes, which describe the image entirely in terms of its own self-similarity.

On Fractal Analogies

Consider the general form of an analogy problem as being **A : B :: C : D**. One can interpret this visually as shown in this figure.



For visual analogy, we can presume each of these analogy elements to be a single image. Some unknown transformation **T** can be said to transform image **A** into image B, and likewise, some unknown transformation **T′** transforms image **C** into the unknown answer image.

The Central Analogy

The central analogy in such a visual problem may then be imagined as requiring that **T** be analogous to **T′**; that is, the answer will be whichever image **D** yields the most analogous transformation. That **T** and **T′** are analogous may be construed as meaning that **T** is in some fashion similar to **T′**. The nature of this similarity may be determined by a number of means, many of which associate visual or geometric features to points in a coordinate space, and compute similarity as a distance metric (Tversky 1977).

As we are interested in discovering which transformation **T′** is most analogous to **T**, we must have at our disposal a set of candidate images to use in the place of **D**. Working from this set of candidates, we may calculate the fractal encoding of the transformation of each candidate image **X** in terms of image **C**. This provides a set of possible transformations, which we shall label **Ω**, from which to seek the most analogous transformation **T'** and thereby find which candidate image was responsible for it.

$\forall$ X $\in$ { candidate images }, Tx ≔ Frac(C,X)
Ω = { T1, T2, T3, T4, … Tn } and T' $\in$ Ω

To determine the transform T′ which is most analogous to transform T from a set of transformations $\Omega \coloneqq \{ T_1, T_2, T_3, T_4, \ldots T_n \}$:

PREPARATORY

Let $\Omega^* \coloneqq \{ T \} \cup \Omega$
Construct a memory M as an empty hash table.
Let F() be a function which generates a set of features.
Let K() be an injective hash function for M.

INDEXING

For each transform $\tau \in \Omega^*$, hash τ in M by:
· Generate a set of features $F(\tau) = \{ f_1, f_2, f_3, \ldots \}$.
· For each feature $f_i \in F(\tau)$, store τ into M, using $K(f_i)$ as a key.

RETRIEVAL

For each transform $T_i \in \Omega$, calculate S_i as the similarity T to T_i by:
· Set $a \leftarrow b \leftarrow c \leftarrow 0$.
· Generate a set of features $F(T_i) \coloneqq \{ f_1, f_2, f_3, \ldots \}$.
· For each feature $f_i \in F(T_i)$:
  • Use $K(f_i)$ as a key to retrieve a set of entries μ from M.
  • If $T \in \mu$, then $a \leftarrow a + 1 \because f_i \in F(T_i) \cap F(T)$.
  • If $T \notin \mu$, then $c \leftarrow c + 1 \because f_i \in F(T_i) - F(T)$.
· Generate a set of features $F(T) \coloneqq \{ f_1, f_2, f_3, \ldots \}$.
· For each feature $f_i \in F(T)$:
  • Use $K(f_i)$ as a key to retrieve a set of entries μ from M.
  • If $T_i \notin \mu$, then $b \leftarrow b + 1 \because f_i \in F(T) - F(T_i)$.
· Calculate S_i from the values a, b, and c:
  $S_i \leftarrow a / (a + \alpha * b + \beta * c)$

Determine $\zeta \leftarrow \max \{ S_1, S_2, S_3, S_4, \ldots S_n \}$

T′ is therefore that transform $T_i \in \Omega$ which corresponds to the maximal similarity ζ, and is deemed the most analogous to transform T.

Implementing Analogy by Recall

Our approach compares each transform in the set **Ω** to the original transform **T** by means of a recall algorithm. This method is divided into stages, as detailed in the algorithm above. We shall describe each stage of the recall algorithm in detail.

Analogy by Recall: Preparatory Stage

Our system uses a feature-based similarity approach to analogy. Consequently, we choose data structures We implement the algorithm by using a hash table as a data structure surrogate for memory **M**. As we will be hashing transformations into **M**, we define two additional operators: **F()**, a method to generate a set of features from a given transformation; and **K()**, an injective hash function which operates solely over the domain of the features.

We made the commitment to a hash table for two reasons beyond that of wishing to use features. First, we note that it is desirous to find some overlap in the features which occur between two transformations, such that a perfect overlap would deem the transformations perfectly analogous. The hash function **K()** may result in hashing multiple transformations to the same feature, and therefore **K()** must operate only upon a given feature, and not take into consideration the transformation which gave rise to that feature. Second, **F()**, the method which generates features from a transformation, must do so in a manner such that each generated feature affords salience, or information content (Tversky 1977).

Analogy by Recall: Indexing Stage

We wish to store each transformation in the hash table memory **M**. The set of possible analogous transformations $\mathbf{\Omega}$ is combined with the original transformation **T** to form a new set $\mathbf{\Omega}^*$. The algorithm iterates over each member $\boldsymbol{\tau}$ of $\mathbf{\Omega}^*$, and from each member calculates a set of features using **F()**. For each feature $\mathbf{f_i}$ in $\mathbf{F(\boldsymbol{\tau})}$, the transformation is indexed into memory **M** as an ordered pair $(\mathbf{K(f_i)}, \boldsymbol{\tau})$. That there likely will be hash collisions at key value $\mathbf{K(f_i)}$ is expected and desired for the retrieval stage.

Analogy by Recall: Retrieval Stage

In the retrieval stage, the algorithm focuses upon determining a measure of similarity between the original transformation **T** and each possible analogous transformation $\mathbf{T_i}$ in $\mathbf{\Omega}$. Our choice of metric reflects similarity as a comparison of the number of fractal features shared between candidate pairs taken in contrast to the joint number of fractal features found in each pair member (Tversky 1977). We desire a metric which is normalized with respect to the number of features under consideration.

In our implementation, the measure of similarity between the target transform **T** and a candidate transform $\mathbf{T_i}$ is calculated using the ratio model (Tversky 1977):

$$\text{similarity}(T,Ti) = a / (a + \alpha * b + \beta * c)$$

where

$$a = \mathcal{F}(T \cap Ti),$$
$$b = \mathcal{F}(T\text{-}Ti),$$
$$c = \mathcal{F}(Ti\text{-}T),$$

and $\mathbf{\mathcal{F}(Y)}$ is a formula which determines the number of features which may be extracted from the set **Y**. The analogy by recall algorithm illustrates how to calculate **a**, **b**, and **c** effectively, using hash table retrieval as a surrogate for distinguishing and counting common and distinct features within the sets $\mathbf{T \cap T_i}$, $\mathbf{T\text{-}T_i}$, and $\mathbf{T_i\text{-}T}$ respectively.

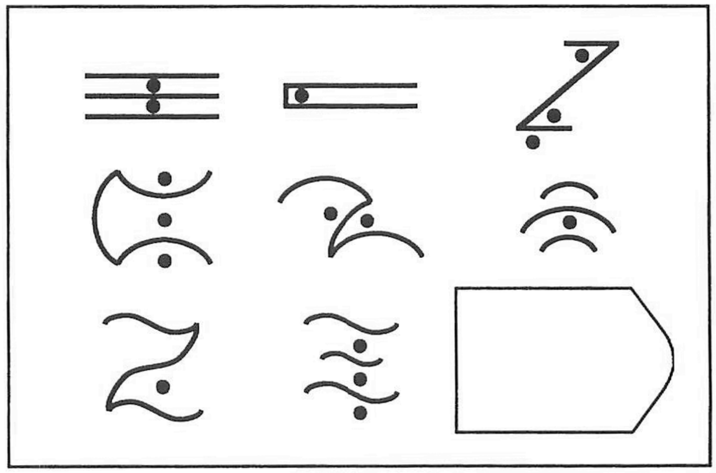
Tversky notes that the ratio model for matching features generalizes several set-theoretical models of similarity proposed in the psychology literature, depending upon which values one chooses for the weights $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. We have found that significant discrimination between candidate answers could be found by setting $\boldsymbol{\alpha \leftarrow \beta \leftarrow 1.0}$, thus favoring features from either transformation equally. As Tversky notes, by equating $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, we ensure that the calculation of similarity is symmetric with respect to the transformations under comparison.

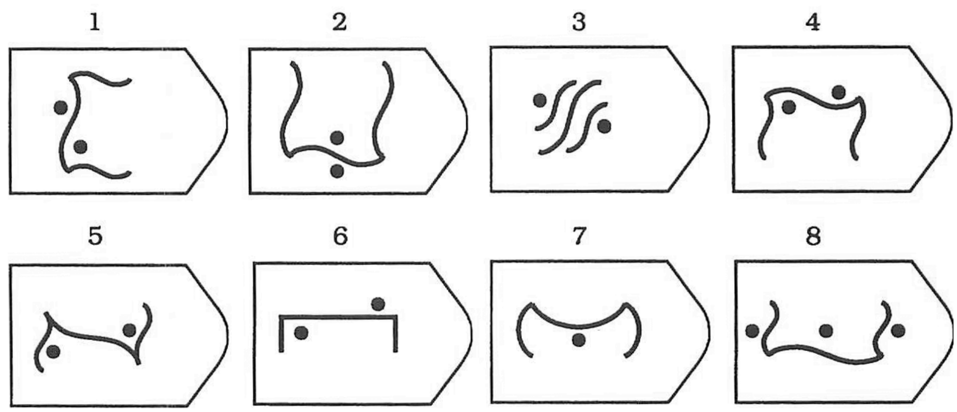
Once the algorithm has calculated the similarity function over all of the candidate transforms, it is a simple matter to determine which transformation has generated the maximal similarity. This transformation, **T'**, is deemed to be the most analogous to the original transformation **T**.

Determining Fractal Features

THE fractal representation of an image is a set of specific fractal codes which compactly describe the geometric alteration and colorization of fragments of the source image that will collage to form the destination image. While it is tempting to treat contiguous subsets of these fractal codes as features, we note that their derivation does not follow strictly Cartesian notions (e.g. adjacent material in the destination might arise from non-adjacent source material). Accordingly, we consider each of these fractal codes independently, and construct candidate fractal features from individual codes.

Each fractal code yields a set of features formed by constructing subsets of its 5-tuple. These features are determined in a fashion to afford position, affine, and colorimetric agnosticism, as well as specificity. In the present implementation, we generate 85 distinct features for each fractal code, by combinations of two, three, and four elements of the 5-tuple.





On The Ravens Progressive Matrices

The Raven's Progressive Matrices (RPM) test paradigm is intended to measure eductive ability, the ability to extract and process information from a novel situation (Raven, Raven, & Court, 2003). Eductive ability stands in contrast to reproductive ability, which is the ability to recall and use previously learned information.

Over the years, different models have proposed various specific mechanisms for solving RPM problems. Hunt (1974) gives a theoretical account of the information processing demands of certain problems from the

Advanced Progressive Matrices (APM), in which he proposes two qualitatively different solution algorithms —"Gestalt," which uses visual operations on analogical representations, and "Analytic," which uses logical operations on conceptual representations.

Carpenter, Just, and Shell (1990) describe a computational model that simulates solving RPM problems using propositional representations. Their model is based on the traditional production system architecture, with a long-term memory containing a set of hand-authored productions and a working memory containing the current state of problem solving (e.g. current goals). Productions are based on the relations among the entities in a RPM problem, for example, the location of the dark component in a row, which might be the top half in the top row of a problem, bottom-half in the bottom row, and so on. They did not test their system on the Standard Progressive Matrices (SPM), but two different versions of their system solved 23 and 32 out of 34 attempted problems on the APM.

Bringsjord and Schimanski (2003) used a theorem-prover to solve selected RPM problems stated in first-order logic, though no results from this effort were reported.

Lovett, Forbus & Usher (2010) describe a model that extracts qualitative spatial representations from visually segmented representations of RPM problem inputs and then uses the analogy technique of structure mapping to find solutions and, where needed to achieve better analogies, to regroup or re-segment the initial inputs to form new problem representations. Again, while visual information from the RPM problems is implicit in the final representations, the structure-mapping engine is applied to these representations without any commitment to the visual nature of the encoded information. This system was tested against sets B through E of the SPM and correctly solved 44 out of 48 attempted problems.

Cirillo and Ström (2010) created a system for solving problems from the SPM that, like that of Lovett et al. (2010), takes as inputs vector graphics representations of test problems and automatically extracts hierarchical propositional problem representations. Then, like the work of Carpenter et al. (1990), the system draws from a set of predefined patterns, derived by the authors, to find the best-fit pattern for a given problem. This system was tested against Sets C through E of the SPM and solved 8, 10, and 10 problems, respectively.

Kunda, McGreggor, & Goel (2010) have developed a model that, operates directly on scanned image inputs from the test. This model uses operations based on mental imagery (rotations, translations, image composition, etc.) to induce image transformations between images in the problem matrix and then predicts an answer image based on the final induced transformation. This model has been tested on all 60 problems from the SPM and correctly solves 35 of these problems.

Finally, Rasmussen and Eliasmith (2011) used a spiking neuron model to induce rules for solving RPM problems. Input images from the test were hand-coded into vectors of propositional attribute-value pairs, and then the spiking neuron model was used to derive transformations among these vectors and abstract over them to induce a general rule transformation for that particular problem. No results are reported for this system.

Tackling Ravens with Fractals

Our interpretation of the Raven's problems follows closely our interpretation of visual analogy problems. The 60 problems from Raven's Standard Progressive Matrices test are organized in five sets, labelled A through E. Each successive set is in general more difficult than the prior set, although problems that human test takers have the most difficulty solving are not necessarily all in set E. Sets A and B are 2x2 matrices of images with six possible answers; the remaining sets are 3x3 matrices of images with eight possible answers. No printed verbal information accompanies the test, though standard administration of the test involves brief oral instructions at the outset.

Reconciling Multiple Analogical Relationships

The analogy by recall algorithm determines the similarity between unique pairs of transforms. However, there may be any number of possible answers to the visual analogy problem: this suggests that many such pairs may be formed.

In our generalized problem, we could presume that not only the horizontal analogical relationship **A : B :: C : ?** is important, but so is the vertical **A : C :: B : ?.** Each relationship compounds the number of similarity comparisons which must be made. We note that while the Raven's SPM problems are similar to the visual analogy problem presented above, the Raven's problems do impose additional analogical constraints. To simultaneously solve horizontal and vertical constraints involves the reconciliation of these multiple analogical relationships.

For each candidate solution, we consider the similarity of each potential analogical relationship as a value upon an axis in a large "relationship space." The dimensionality of this space is determined by the problem at hand.

To specify the overall fit of a candidate solution, we construct a vector in this multidimensional relationship space and determine its length, using a Euclidean distance formula. The longer the vector, the more similar two members are; the shorter the vector, the more dissimilar two members are.

Ambiguity

It would seem a trivial matter to identify which one among the candidate images has the highest similarity score. As it turns out, this is not yet sufficient for solving the problem, as ambiguity may be present.

Similarity scores for the candidate images may vary widely. If the score for any candidate is unambiguously larger than that of any other candidate, then the candidate is deemed the most similar. We calculate the deviation of each similarity score from the average of all such scores, and use confidence intervals as a means for indicating ambiguity. In our present implementation, we use a confidence level of 90%. We note that ambiguity also is indicated if more than one candidate has a high confidence ranking.

Refinement strategy

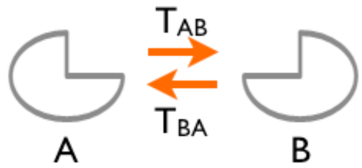
If the algorithm's output for a particular problem is ambiguous, then there are two available remedial mechanisms: to modify the grouping such that larger sets of images are considered simultaneously (from pairs to triplets); or to recalculate the fractal representations using a finer partitioning.

Mutuality

The analogical relationship between source and destination images may be seen as mutual. That is, the source is to the destination as the destination is to the source. However, the fractal representation which entails encoding is decidedly directional (e.g. from the source to the destination). To capture the bidirectional, mutual nature of the analogy between source and destination, we introduce the notion of a mutual fractal representation.

Let us call the representation of the fractal transformation from image **A** to image **B** as $\mathbf{T_{AB}}$, as shown.

Correspondingly, we would label the inverse representation as $\mathbf{T_{BA}}$.



We shall define the mutual analogical relationship between **A** and **B** by the symbol $\mathbf{M_{AB}}$, given by this equation:

$$M_{AB} = T_{AB} \cup T_{BA}$$

By exploiting the set-theoretic nature of $\mathbf{T_{AB}}$ and $\mathbf{T_{BA}}$ to express $\mathbf{M_{AB}}$ as a union, we afford the mutual analogical representation the complete expressivity and utility of the fractal representation.

Extending Mutuality

The mutual fractal representation of the pairings may be employed to determine similar mutual representations of triplets of images. For example, the mutual fractal relationship between three images **A**, **B**, and **C** may be represented by the symbol $\mathbf{M_{ABC}}$, and defined by this equation:

$$M_{ABC} = M_{AB} \cup M_{AC} \cup M_{BC}$$

Fractal Refinement

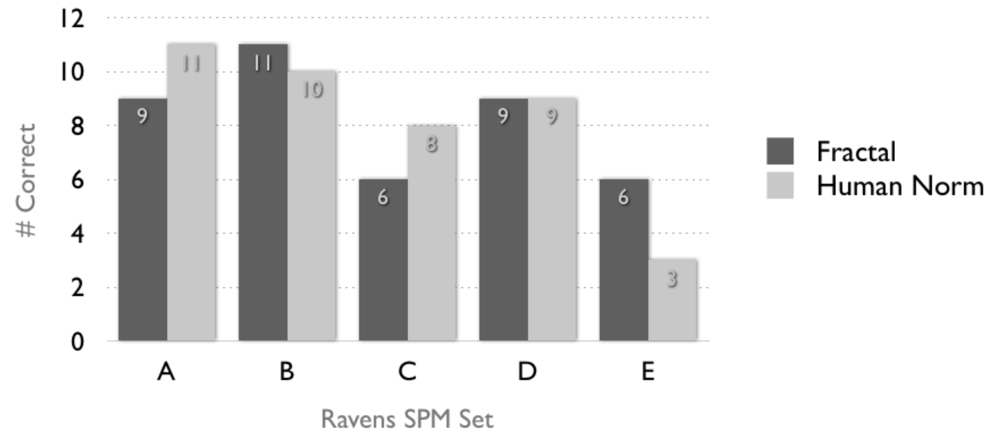
At present, our implementation attempts bumping up the images considered simultaneously as a first measure, although this is only practical on SPM problems in sets C through E, as they are 3x3 matrices, and afford triple groupings horizontally and vertically. If the scoring remains ambiguous after reaching a grouping based upon triplets, then we consider that the initial representation level was too coarse, and rerun the algorithm using ever finer partitions for the mutual fractal representation, starting once more with pairs. At last, if after altering our considerations of groupings and examining the images at the finest level of resolution the scores prove inconclusive, the algorithm quits, choosing the answer in which it found the highest confidence.

Results

We tested our technique on all 60 problems of the Raven's Standard Progressive Matrices test. To create inputs for the fractal algorithm, each page from the SPM test booklet was scanned, and the resulting greyscale images were rotated to roughly correct for page alignment issues. Then, the images were sliced up to create separate image files for each entry in the problem matrix and for each answer choice. These separate images were the inputs to the technique for each problem. No further image processing or cleanup was performed, despite the presence of numerous pixel-level artifacts and minor alignment issues. Additionally, the fractal algorithm attempted to solve each SPM problem independently: no information was carried over from problem to problem.

Fractal Performance

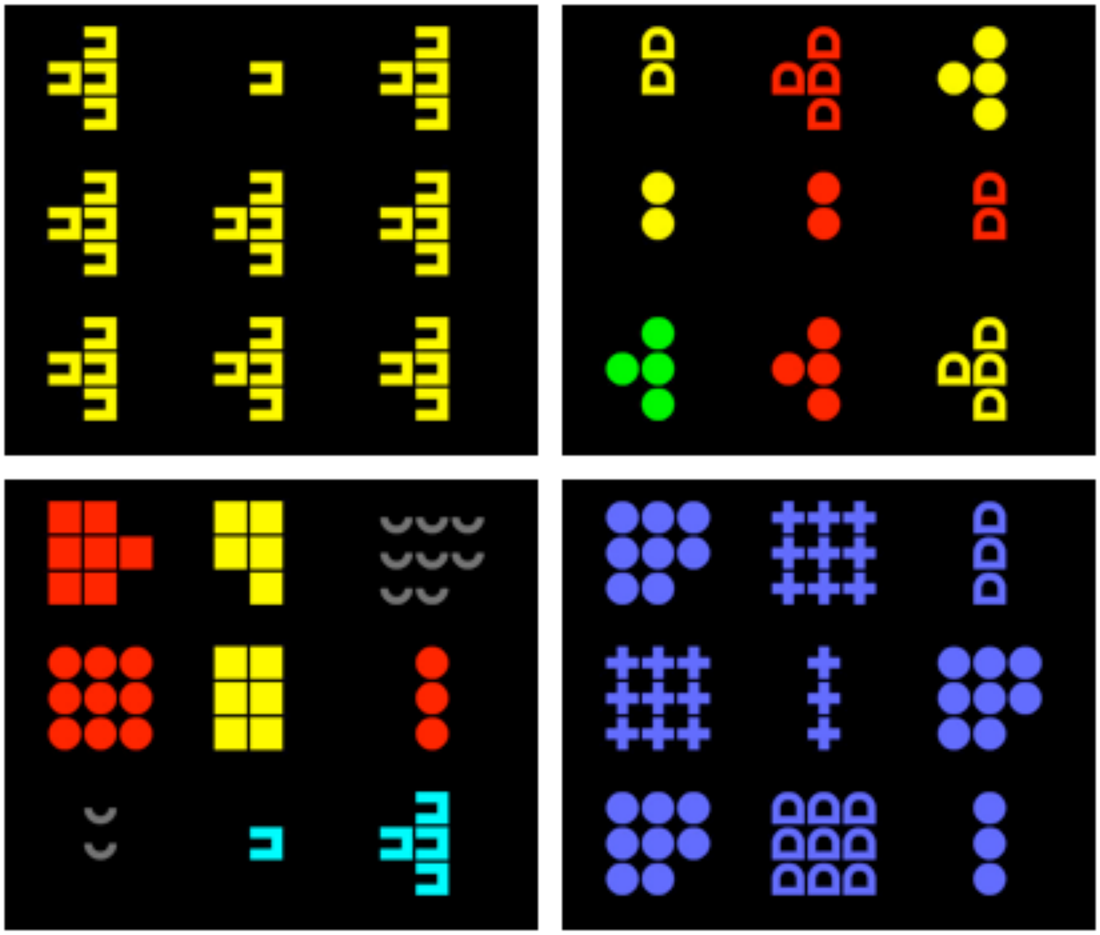
There are three main assessments that can be made following the administration of the SPM to an individual: the total score, which is given simply as the number of correct answers; an estimate of consistency, which is obtained by comparing the given score distribution to the expected distribution for that particular total score; and the percentile range into which the score falls, for a given age and nationality (Raven, Raven, & Court 1998).



The total score obtained by the fractal algorithm was 41 correct out of 60 problems. The score breakdown by set, along with the expected score composition for a total score of 41 are shown in Figure 5. A score is "consistent" if the difference between the actual score and the expected score for any given set is no more than +/-2 (Raven, Raven, & Court 1998). Inconsistent scores may result from test takers not understanding the test instructions, randomly guessing, or trying to choose incorrect answers to artificially lower the total score, for example.

The score differences for the fractal algorithm on each set were no more than +/-2, with the exception of set E. For a human test-taker, this score distribution generally would indicate that the test results do provide a valid measure of the individual's general intellectual capacity. This score pattern illustrates that the results achieved by the algorithm fall well within typical human norms on the SPM for sets A-D. The algorithm's performance on set E exceeds expected human norms.

Finally, the total score can be compared to age-group and national norms to determine percentile rankings. Using norms from the United States, we see that a total score of 41 corresponds to the 95th percentile for children about 9 years old, the 50th percentile for children around 13 years old, and the 10th percentile for children older than 17.5 years old (Raven, Raven, & Court 1998).



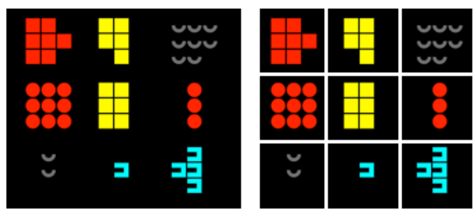
On The Odd One Out

The Odd One Out test of intelligence (Hampshire 2010) consists of 3x3 matrix reasoning problems organized in 20 levels of difficulty. In the test, a participant must decide which of the nine abstract figures in the matrix does not belong (the so-called "Odd One Out"). Figure 3 shows a sampling of the problems, illustrating the nature of the task, and several levels of complexity.

From the computational perspective, one drawback of computationally modeling a visual analogy task such as the Raven's test is that the algorithm for generating the problems on the test is not known; human examiners generate the test problems based on historical and empirical data. In contrast, problems on the Odd One Out test are generated using a complex set of algorithms (Hampshire 2010). Thus, many tens of thousands of novel problems may be generated.

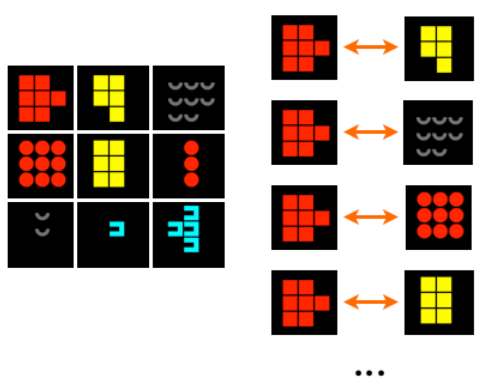
Finding the Odd One Out, Fractally

We now present our algorithm for tackling the Odd One Out problem, using the mutual fractal representation as a basis for visual reasoning. The algorithm consists of three phases: segmentation, representation, and reasoning.



The segmentation phase.

First, we must segment the problem image P into its nine constituent subimages, I_1 through I_9. In the present implementation, the problems are given as a 478x405 pixel JPEG image, in the RGB color space. The subimages are arrayed in a 3x3 grid within the problem image. At this resolution, we have found that each subimage fits well within a 96x96 pixel image, as may be seen in Figure 4.



The representation phase.

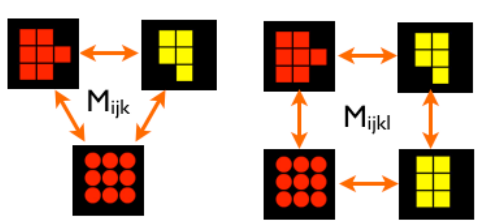
We next must transform the problem into the domain of fractal representations. Given the nine subimages, we group subimages into pairs, such that each subimage is paired once with the other eight subimages. Thus, we form 36 distinct pairings. We then calculate the mutual fractal representation M_{ij} for each pair of subimages I_i and I_j. We determine the fractal transformation from I_i to I_j in the manner described in (McGreggor et al. 2011), then form the union of the sets of codes from the forward and backward fractal transformation to construct M_{ij}.

The block partitioning we use initially is identical to the largest possible block size (in this case, 96x96), but subsequent recalculation of M_{ij} may be necessary using finer block partitioning (as proscribed in the reasoning phase). In the present implementation, we conduct the finer partitioning by uniform subdivision of the images into block sizes of 48x48, 24x24, 12x12, 6x6, and 3x3.

Extended Mutuality

At this phase, we note that the mutual fractal representation of the pairings may be employed to determine similar mutual representations of triplets or quadruplets of images. These subsequent representations may be required by the reasoning phase. As a notational convention, we construct these additional representations for triplets (M_{ijk}) and quadruplets (M_{ijkl}) in this manner:

$$M_{ijk} = M_{ij} \cup M_{jk} \cup M_{ik}$$
$$M_{ijkl} = M_{ijk} \cup M_{ikl} \cup M_{jkl} \cup M_{ijl}$$



The reasoning phase.

We shall determine the odd one out solely from the mutual fractal representations, without reference or consideration to the original imagery. We start by considering groupings of representations, beginning with pairings, and, if necessary, advance to consider other groupings.

Reconciling Multiple Analogical Relationships

For a chosen set of groupings, G, we must determine how similar each member is to each of its fellow members. We first derive the features present in each member, as described above, and then calculate a measure of similarity as a comparison of the number of fractal features shared between each pair member (Tversky 1977).

We desire a metric of similarity which is normalized with respect to the number of features under consideration, and where the value 0.0 means entirely dissimilar and the value 1.0 means entirely similar. Accordingly, in our present implementation, we use the ratio model of similarity as described in (Tversky 1977). According to the ratio model, the measure of similarity S between two representations A and B is calculated thusly:

$$S(A,B) = f(A \cap B) \ / \ [f(A \cap B) + \alpha f(A\text{-}B) + \beta f(B\text{-}A)]$$

where $f(X)$ is the number of features in the set X.

Tversky notes that the ratio model for matching features generalizes several set-theoretical models of similarity proposed in the psychology literature, depending upon which values one chooses for the weights α and β. To favor features from either image equally, we have chosen to set $\alpha = \beta = 1$.

Relationship Space

As we perform this calculation for each pair A and B taken from the grouping G, we determine for each member of G a set of similarity values. We consider the similarity of each analogical relationship as a value upon an axis in a large “relationship space” whose dimensionality is determined by the size of the grouping: for pairings, the space is 36 dimensional; for triplets, the space is 84 dimensional; for quadruplets, the space is 126 dimensional.

Treating Maximal Similarity as Distance

To arrive at a scalar similarity score for each member of the group G, we construct a vector in this multidimensional relationship space and determine its length, using a Euclidean distance formula. The longer the vector, the more similar two members are; the shorter the vector, the more dissimilar two members are. As the Odd One Out problem seeks to determine, literally, "the odd one out," we seek to find the shortest vector, as an indicator of dissimilarity.

Distribution of Similarity

We have determined a score for the grouping G, but have not yet arrived at individual scores for the subimages. To determine the subimage scoring, we distribute the similarity equally among the participating subimages. For each of the nine subimages, a score is generated which is proportional to its participation in the similarity of the grouping's similarity vectors. If a subimage is one of the two images in a pairing, as an example, then the subimage's similarity score receives one half of the pairing's calculated similarity score.

Once all of the similarity scores of the grouping have been distributed to the subimages, the similarity score for each subimage is known. It is then a trivial matter to identify which one among the subimages has the lowest similarity score. As it turns out, this is not yet sufficient for solving the problem, as ambiguity may be present.

Ambiguity

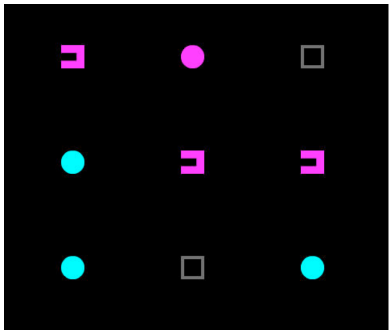
Similarity scores for the subimages may vary widely. If the score for any subimage is unambiguously smaller than that of any other subimage, then the subimage is deemed "the odd one out." By unambiguous, we mean that there is no more than one score which is less than ε, which we may vary as a tuning mechanism for the algorithm, and which we see as a useful yet coarse approximation of the boundary between the similar and the dissimilar in feature space. In practice, we calculate the deviation of each similarity measure from the average of all such measures, and use confidence intervals (as calculated from the standard deviation) as a means for indicating ambiguity.

Refinement strategy

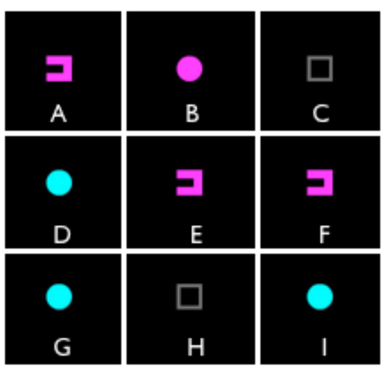
However, if the scoring is inconclusive, then there are two readily available mechanisms at the algorithm's disposal: to modify the grouping such that larger sets of subimages are considered simultaneously (from pairs to triplets, or from triplets to quadruplets), or to recalculate the fractal representations using a finer partitioning. In our present implementation, we attempt bumping up the elements considered simultaneously as a first measure. If after reaching a grouping based upon quadruplets the scoring remains inconclusive, then we consider that the initial representation level was too coarse, and rerun the algorithm using ever finer partitions for the mutual fractal representation. If, after altering our considerations of groupings and examining the images at the finest level of resolution the scores prove inconclusive, the algorithm quits, leaving the answer unknown.

Example

We now present an example of the algorithm, selected for its illustrative power, and not for its difficulty. We do note that the answer to this particular problem may be at once obvious to the reader.



The algorithm begins by segmenting the image into the nine subimages. For convenience, let us label the images A through I, as shown in Figure 8. Once segmented, fractal representations are formed for each possible pairing of the subimages, for a total of 36 distinct representations. The initial partitioning of the subimages for fractal encoding shall be at the coarsest possible level, 96x96 pixels.



In this example, it is quite clear to the reader that there are pairings which are identical (e.g. {A,E}, {E, F}, {A, F}, {C, H}, {D, G}, {D, I}, and {G, I}). The fractal representation of each of these pairings, at this coarsest level of partitioning (96x96) will yield the Identity transformation, with zero photometric correction. Thus the similarity between these particular transform pairs will be 1.0. These pairings we shall deem therefore to be perfectly analogous. However, not all of the representations will be similar. For example, the pairing of subimage C to any image other than subimage H will result in a substantially different fractal encoding than the {C,H} pairing.

For each subimage, we calculate the similarities of all eight possible pairings of that subimage against all other unique pairings. We next construct a similarity vector in 36-space for each pairing.

We derive the length of the 36-tuple similarity vector, normalize the result, and distribute this value to each of the subimages involved in the pairing by summing. In this example, the length of the similarity vector for the pairing {A,B} is found to be 4.55, we divide by 6 (the length of a 36-tuple with all entries 1.0), for a value of 0.7583. This value is added to the current summation for subimages A and B. At the close of this process, each subimage will have a score, representing the distributed similarity scores for all of the pairings in which it played a part.

The algorithm then examines the set of scores for all of the subimage, looking for ambiguity. Our present implementation defines ambiguity as the data having more than one item which deviates from the mean by a value greater than the standard deviation of the data. If this holds true, then the result is deemed ambiguous.

96x96		
0.727	0.727	$\boxed{0.674}$
0.716	0.727	0.727
0.716	$\boxed{0.674}$	0.716
μ : 0.711, σ : 0.022		

48x48		
0.534	$\boxed{0.413}$	0.468
0.510	0.534	0.534
0.510	0.468	0.510
μ : 0.499, σ : 0.041		

Table 1 illustrates the ambiguity found in using a 96x96 partitioning of the subimages, with two values having a deviation of 1.713σ. Thus, the algorithm must proceed to a finer partitioning in order to produce an unambiguous answer. Using a 48x48 partitioning, produces a single unambiguous result, with a deviation of 2.117σ. Accordingly, subimage B is selected as the Odd One Out.

Results, Preliminary Analysis and Discussion

We have run our algorithm against 2,976 problems of the Odd One Out. These problems were randomly selected from a span of difficulty from the very easiest (level one) up to the most difficult (level 20). The example problem presented in Figure 7 is a level 11 problem.

We restricted the algorithm to attend only to pairings of subimages, and to progress from an initial partitioning of 96x96 blocks (essentially, the entire subimage) to no further refinement of partitioning than 6x6. We made these restrictions in order to fully exercise the strategic shifting in partitioning, to assess the similarity calculations, and to judge the effect of mutuality, at a tradeoff in execution time. The results are presented in the table below.

Level	Total	Correct	6x6	12x12	24x24	48x48	96x96
1	148	147	0	0	0	1	0
2	148	135	1	0	1	3	8
3	147	119	3	1	3	5	16
4	149	141	0	0	1	2	5
5	149	88	13	2	9	11	26
6	149	97	17	3	0	14	18
7	149	100	9	0	4	16	20
8	149	88	13	3	3	13	29
9	149	114	7	3	8	10	7
10	149	125	9	2	0	7	6
11	149	115	11	1	4	11	7
12	149	123	5	5	4	4	8
13	149	36	28	2	12	28	43
14	149	38	17	3	14	24	53
15	149	36	26	6	10	22	49
16	149	34	23	3	8	32	49
17	149	22	26	9	24	38	30
18	149	28	24	10	16	41	30
19	149	31	23	10	25	27	33
20	149	30	25	12	11	41	30
Total	**2976**	**1647**	**280**	**75**	**157**	**350**	**467**

We note that there are quite clear degrees of performance variation generally grouped according to sets of levels (levels 1-4, 5-8, 9-12, 13-16, and 17-20). This is consistent with the (unknown both to us and to the algorithm) knowledge that the problems at these levels were generated using varying rules. Our algorithm at present does not carry forward information between its execution of each problem, let alone between levels of problems. However, that the output illustrates such a strong degree of performance shift provides a further research opportunity in the areas of reflection, abstraction and meta-reasoning, in the context of the original fractal representations.

The rightmost five columns of the results data provide a breakdown of errors made at differing partitioning levels. Immediately the reader will note that the majority of errors occur when the algorithm stops at quite high levels of partitioning (96x96 or 48x48). We interpret this as strong evidence that there exists levels-of-detail (or partitioning) which are too gross to allow for certainty in reasoning. Indeed, the amount of data upon which decisions are made at these levels are three orders of magnitude less than that which the finest partitioning affords (roughly 100 features at 96x96 versus more than 107,000 features at 6x6). We find an opportunity for a refinement of the algorithm to assess its certainty (and therefore, its halting) based upon a naturally emergent artifact of the representation.

A temptation might be to reverse the partitioning process, beginning at the finest partition (6x6) and progress upward until ambiguity appears due to insufficient level of detail. In an earlier test of this notion, using a random sampling of problems across a span of difficulty levels, we found that ambiguity existed at both small and large levels of detail; that is, that ambiguity exists at either too fine or too large a level of detail, and that an unambiguous answer arose once some sufficiency in level of detail was realized. It is important to note that the sufficient level of detail was discoverable by the algorithm, emerging from the features derived from the fractal representation.

The errors which occurred at the finest level of partitioning (6x6) are caused not due to the algorithm reaching an incorrect unambiguous answer (though this is so in a few cases) but rather that the algorithm was unable to reach a sufficiently convincing or unambiguous answer. As we noted, these results are based upon calculations involving considering shifts in partitioning only, using pairwise comparisons of subimages. Thus, there appear to be Odd One Out problems for which considering pairs of subimages shall prove inconclusive (that is, at all available levels of detail, the results will be found to be ambiguous). It is this set of problems which we believe implies that a shift in grouping (from pairs to triplets, or from triplets to quadruplets) must be undertaken to reach an unambiguous answer.

Papers, Posters, and Publications

Aspects of my research have been published or presented in various forms. Here is a list of those publications.

Publications in progress

- Kunda, M., McGreggor K., and Goel, A. Visual Reasoning on the Raven's Advanced Progressive Matrices Test. Submitted to the 34th Annual Meeting of the Cognitive Science Society, August 2012.
- McGreggor, K., and Goel, A. A Fractal Strategy for Noticing Novelty. Submitted to the 26th National Conference on AI (AAAI-2012), Cognitive Systems Track, August 2012, Toronto, Ontario.
- McGreggor, K., and Goel, A. Froids: An Experiment in Fractal Perception. Submitted to the Third International Conference on Computational Creativity, May 2012, Dublin, Ireland.
- McGreggor, K., Kunda, M., and Goel, A. Fractals and Ravens. Submitted to Artificial Intelligence Journal, forthcoming.

Accepted publications

- McGreggor, K., and Goel, A. Fractally Finding the Odd One Out: An Analogical Strategy For Noticing Novelty. Fall AAAI Symposium on Cognitive Systems, November 2011, Arlington, VA. (Poster)
- McGreggor, K., and Goel, A. Finding The Odd One Out: A Fractal Analogical Approach. 8th ACM Conference on Creativity and Cognition, November 2011, Atlanta, GA.
- Kunda, M., McGreggor, K., and Goel, A. Two Visual Strategies for Solving the Raven's Progressive Matrices Intelligence Test. 25th National Conference on AI (AAAI-2011), New Scientific and Technical Advances in Research (NECTAR) Track , August 2011, San Francisco, CA.
- McGreggor, K., Kunda, M., and Goel, A. Fractal Analogies: Preliminary Results from the Raven's Test of Intelligence. Second International Conference on Computational Creativity, April 2011, Mexico City, Mexico.
- Kunda, M., McGreggor, K., and Goel, A. Taking A Look (Literally!) At The Raven's Intelligence Test: Two Visual Solution Strategies. 32nd Annual Meeting of the Cognitive Science Society, August 2010.
- McGreggor, K., Kunda, M., and Goel, A. A Fractal Analogy Approach To The Raven's Test Of Intelligence. Workshop on Visual

Representations and Reasoning, 24th National Conference on AI (AAAI-2010), July 2010, Atlanta, GA.

- Kunda, M., McGreggor, K., & Goel, A. Can the Raven's Progressive Matrices intelligence test be solved by thinking in pictures? Nineth Annual International Meeting For Autism Research (IMFAR), May 2010, Philadelphia, PA. (Poster)
- McGreggor, K., Kunda, M., and Goel, A. A Fractal Approach Towards Visual Analogy. First International Conference on Computational Creativity, January 2010, Lisbon, Portugal.
- Kunda, M., McGreggor, K., and Goel, A. Addressing The Raven's Progressive Matrices Test Of "General" Intelligence. Fall AAAI Symposium on Multimodal Representations, November 2009, Arlington, VA.

References

Here are some selected references, works either directly referred to by this proposal, or ones that inspired my research.

- Barnsley, M. F., & Hurd, L. P. 1992. Fractal Image Compression. Boston, MA: A.K. Peters.
- Bringsjord, S., & Schimanski, B. 2003. What is artificial intelligence? Psychometric AI as an answer. IJCAI, 18, 887–893.
- Carpenter, P., Just, M., & Shell, P. 1990. What one intelligence test measures: a theoretical account of the processing in the Raven Progressive Matrices Test. Psychol. Review, 97, 404-31.
- Cirillo, S., & Ström, V. (2010). An Anthropomorphic Solver for Raven's Progressive Matrices (No. 2010:096). Goteborg, Sweden: Chalmers University of Technology.
- Davis, R., Shrobe H., and Szolovits, P. "What is a Knowledge Representation?" AI Magazine 14.1 (1993):17-33.
- Frisby, J.P., and Stone, J.V. Seeing: the Computational Approach to Biological Vision, 2nd edition. Cambridge: MIT Press, 2010.
- Hunt, E. 1974. Quote the raven? Nevermore! In L. Gregg (Ed.), Knowledge and cognition (pp. 129–158). Hillsdale, NJ: Erlbaum.
- James, W. The Principles of Psychology, 2 vols. (1890) Dover Publications, 1950.
- Kunda, M., McGreggor, K., & Goel, A. (2010) Taking a Look (Literally!) at the Raven's Intelligence Test: Two Visual Solution Strategies. In Proc. 32nd Annual Meeting of the Cognitive Science Society, Portland.
- Lovett, A., Forbus, K., & Usher, J. 2010. A structure-mapping model of Raven's Progressive Matrices. In Proceedings of the 32nd Annual Meeting of the Cognitive Science Society, Portland, OR.
- Mandelbrot, B. 1982. The fractal geometry of nature. San Francisco: W.H. Freeman.
- Marr, D. Vision. A Computational Investigation into the Human Representation and Processing of Visual Information. W.H. Freeman and Company, 1982
- Poggio, T. "Marr's Approach to Vision." AI Memo 645, Massachusetts Institute of Technology, Artificial Intelligence Laboratory, August 1981.

- Posner, M.I., and Petersen, S.E. "The Attention System of the Human Brain." Annual Review of Neuroscience 13 (1990): 25-42. Evans, T. G. 1968. A program for the solution of a class of geometric-analogy intelligence-test questions. Semantic Information Processing, pp. 271–353, MIT Press.
- Raven, J., Raven, J.C., & Court, J. H. Manual for Raven's Progressive Matrices and Vocabulary Scales, General Overview. San Antonio, TX: Harcourt Assessment. (1998)
- Rasmussen, D., & Eliasmith, C. (2011). A neural model of rule generation in inductive reasoning. Topics in Cognitive Science, 3(1), 140-153.
- Treisman, A. "Preattentive Processing in Vision." Computer Vision, Graphics, and Image Processing 31 (1985): 156-177.
- Treisman, A.M., and Gelade, G. "A Feature-Integration Theory of Attention." Cognitive Psychology 12 (1980): 97-136.
- Tversky, A. 1977. Features of similarity. Psychological Review 84 (4): 327-352.
- Wolfe, J.M. "Visual Search." Attention, H. Pashler (Ed.), London, UK: University College London Press, 1998.